"Read my lips" – the commitment value of empty promises under loss-aversion.*

John Hassler[†], **José V. Rodriguez Mora** [‡]

May, 2006

Abstract

Politics seems full of cases where promises that appear empty nevertheless affect later political outcomes, therefore providing an *ex-ante* commitment value. Similarly, political discusion is often asymmetrically focused on individuals who might "loose" from policy changes relative to *status quo*. This paper argues that such observations easily can be explained by introducing reference dependent utility – loss aversion – into a standard political economy model. While experimental evidence shows substantial support for this type of utility, it has not been used in dynamic political economy models before. We show that loss-aversion can provide a basis for (limited) commitment. This can help overcome time inconsistency in capital taxation and also explain why political promises can affect future outcomes. We also show that the willingness to use this mechanism to constrain future policy hinges on the possibiity that future policy makers might not be equally forward looking. Thus, future political myopia breeds current forward looking behavior and in some circumstances can this can create oscillatory policies.

*We thank seminar participants at IIES, Hebrew University, Fudan University and at the ESSIM/CEPR conference for comments and suggestions.

[†]John Hassler: IIES and CEPR; e-mail: John@hassler.se.

[‡]José V. Rodriguez Mora: University of Southampton, Universitat Pompeu Fabra, and CEPR; e-mail:sevimora@upf.es.

1 Introduction

"Read my lips – no new taxes"

Presidential candidate George Bush, Republican party national convention, August 18, 1988

(speech writer: Peggy Noonan)

At least since the seminal work by (Kydland and Prescott 1977) time inconsistency and the difficulty to commit to optimal policies has been recognized as a problem inherent in many aspects of policy making. (Kydland and Prescott 1977) discuss several examples of where the ability or inability to commit is of keys importance for the chosen policy and for welfare. The examples include; insurance against natural disasters, patent policy, capital income taxation and monetary policy. The ideas in this seminal paper has lead to a very large literature on the importance of commitment.[1]

A key implication of the previous literature is that the welfare loss of no ability to commit future policy can be high. More specifically, Markov equilibria in models of time-inconsistencies in the vein of (Kydland and Prescott 1977) can be very different from the commitment solution, both in terms of observables and welfare. Empirically, however, the Markov equilibrium does often not seem to be a good description of reality – capital is not expropriated and inflation rates are moderate, indicating that policy makers somehow manage to commit to policies (to rules) rather than fall for the temptation of discretion.

There are several classical explanations of how policy makers manage to overcome the time inconsistency problem. One is that power can be delegated to a person with preferences different from society or the decisive agent. The seminal paper here is (Rogoff 1985). In a sense, however, this explanation begs the question how commitment actually is achieved. The second explanation is to consider infinite horizon

[1] A completely non-exhaustive list can begin with *(Barro and Gordon 1983)*,

games with multiple equilibria in non-Markovian strategies. By the folk theorem, the commitment solution or at least something close to it can be achieved by the threat of reverting to a bad equilibrium for an extended period of time unless discount rates are too high. The trigger strategies used to achieve the commitment solution in a game with a sequence of different voters require a substantive amount of intergenerational coordination sometimes labeled a "social contract". In particular, to prevent deviations, voters must be confident that future voters will punish current deviations from the social contract by coordinating on "bad" equilibrium without attempting to re-negotiate the contract. Additionally there is some experimental evidence in political games suggesting that even in the much more simple lab-environment, trigger strategies are too difficult to form a basis for a better equilibrium than the Markovian (Cabrales, Nagel and Mora 2006). In this paper we present a new and alternative explanation to how an equilibrium resembling the commitment solution, i.e., involving moderate taxation on sunk investments, can be sustained. In contrast to alternative explanations, equilibria in our model require no intertemporal coordination being instead Markovian in nature.

Our idea is simply that people dislike being fooled. Suppose an individual believes she has been promised to keep some return from an investments she undertakes. If the promise later is broken, the individual is likely to feel a utility loss over and above the direct monetary loss. To prevent this utility loss, an individual may be willing to incur a cost in other dimensions, or example by engaging in costly punishments of political candidates that try to fool them. To model this formally, we build on prospect theory, with loss aversion as formalized by (Kahneman and Tversky 1991)). It is now well known that standard expected utility theory often fails to explain observed individual behavior and that prospect theory can provide better explanations. (Bateman, Munro, Rhodes, Starmer and Sugden 1997) provides experimental evidence on individual valuation of private goods. (Bowman, Minehart and Rabin 1999) show that the behavior of aggregate consumption deviates from the predictions of standard expected utility in a way consistent with prospect theory. (Quattrone and Tversky 1988) provide evidence on systematic deviations from standard expected utility in voting-like experiments. (Kahneman and Tversky 1991) also argues that the empirical finding that the political incumbent advantage is stronger in good times is evidence in favor or voters having reference dependent utility.[2] The accumulating evidence on the empirical relevance of reference dependent utility provides a strong argument for an explorative analysis of the consequence of including such utility in models of political economy. To our knowledge, however, we are the first to introduce loss aversion in dynamic political economy models.

We argue that prospect theory can help understanding important issues in politics that are difficult to account for in standard theory. The focus in this paper is on commitment, although other applications, like status *quo bias* and explanations for why some issues become salient in political campaigning and other not, are quite apparent. We use a very simple dynamic politico-economic model of investment. In every period, there is a stock of sunk investments. Under the assumption that the shadow value of public funds is large enough, there is temptation to use this *ex-post* non-distortionary source of revenue. Without loss-aversion, the only equilibrium in a finite horizon game is one with 100% taxation but since this is foreseen by rational investors, no investment will be undertaken. However, loss aversion introduces a possibility to affect future policy outcomes, creating a strategic link between consecutive policy makers also in the Markov case. Specifically, loss-aversion creates a form of limited commitment – if individuals expect to get back something from their investments, it becomes more costly *ex-post,* not to give it.

The dynamics of the equilibrium will depend on the dynamics of reference point formation. We will consider two polar cases. In the first, we assume that the reference point is forward looking and deter-

[2] The argument is based on the fact that under loss-avesion, individuals are risk-loving in losses. Then, it is better to take a chance with a more risky outsider of equal expected quality if the incumbent gives something below reference for sure.

mined by rational expectations about future consumptions levels. A similar assumption on reference point formation is done in (Köszegi and Rabin 2006). In the opposite case, we assume that the reference point is backward looking and, in particular, (partly) determined by past taxation. Dynamics in the two cases will differ, but interestingly, the steady state is identical.

In the forward looking case, we assume that the reference point of an individual can be anything that is consistent with rational expectations, i.e., that is in the set of equilibrium outcomes of the political game. Of course, this implies the possibility of a large set of self-fulfilling equilibria. However, we will allow a mechanism whereby politicians can affect future reference points and we believe that we can interpret this as political promises. Governments or political candidates can make promises about their future intentions. If such promises are incompatible with all future equilibria, they will not be believed and have no impact on the outcome. However, if a promise is consistent with a future equilibrium, we allow individual reference point formation to be coordinated on the promise, affecting future policy outcomes. Empirically, it is an understatement to say that making promises is part of political campaigning. It seems as these promises are far from meaningless but are an important part of policy making, both before elections and after. The right promise can make a candidate win but can also haunt the winning candidate afterwards if it becomes difficult or impossible to fulfill the promise. The strong effect of promises is, however, difficult to understand in standard models since they often seem not be supported by any clear commitment mechanism.

In the case of backward looking reference points, a reduction in current tax rates increases reference point consumption and will tend to restrain future taxation. However, this does not necessarily lead benevolent policy makers to reduce taxation today. In fact, if it were an equilibrium policy to restrain taxation today, such a policy must be an equilibrium also next period. Then, it would be better to tax away any sunk investment today, and wait for next period's policy makers to restrain taxation. In other words, policy makers would have an incentive to procastinate and delay to the following period the cost of committing to a policy. However, since a cut in taxation can beneficially reduce taxation next period, 100% taxation in every period is not an equilibrium either. In fact, the possibility that future policy makers *will not* restrain taxation voluntarily, creates an incentive for policy makers to restrain it today *and vice-versa*. In a sense that will be clear below, forward looking behavior of current policy makers requires the possibility of future myopia. This strategic substitutability implies oscillations in political strategies when these are required to be pure. Allowing, mixed strategies randomization between forward looking and myopic strategies constitutes a Markov equilibrium with limited taxation in the infinite horizon version of the model. This equilibrium critically hinges on the possibility that future tax rates might be set without consideration of the future.

As already noted, we set up our model so that 100% taxation is the only finite horizon equilibrium. Of course, a large set of other equilibria exists if more elaborate trigger strategies are allowed in the infinite horizon case. Such strategies would require a large, perhaps unreasonably large, amount of coordination among voters and across generations to sustain equilibria with lower taxes. Nevertheless a natural, and very important question is whether the hypothesis of loss-aversion can be empirically distinguished from trigger strategies. We argue that is, at least in principle, is possible. To sustain "good" trigger strategy equilibria, it is key that the subjective discount rate is not too large. Under loss-aversion, the rate of discounting is irrelevant for the outcome. Furthermore, trigger strategy equilibria are sustained by threats of reversion to a worse equilibria and the "best" equilibrium is sustained by the treat of reverting to the worst equilibrium for ever. Under loss aversion, these punishment phases are non-existent. It is of course not immediate that this difference is possible to use to distinguish loss-aversion from trigger strategies since punishments are out of equilibrium. However, in some stochastic games with incomplete information, punishment phases must occur with positive probability. We think that this may provide a way to test our theory.

The paper is organized as follows: In section 2, we present the model, including the economic environment and preferences. Section 3 describes the determination if taxes. Section 4 and 5 describes respectively, equilibria without and with loss aversion. Section 6 provides some concluding remarks.

2 The model

Our model economy has a standard time-inconsistency problem due to the possibility to tax sunk investments at no *ex-post* distortionary cost. It has a two-period OLG structure – a unitary mass of identical, atomistic and non-altruistic individuals who live for two periods and, consequently, there is in each period a cohort of young and old individuals alive. In the first period of life, the individuals have access to an investment project and individually choose an investment level i_t but are, for simplicity, assumed not to consume. The investment is costly, incurring an immediate utility cost i_t^2. In the second period of life, the individual consumes a private and a public good, denoted c_{t+1} and G_{t+1} respectively. We normalize the gross expected return on the investment to unity. Given a tax-rate τ_t, an individual born in period $t-1$ solves

$$\begin{aligned} U_t &= \max_{c_{t+1}, i_t} -\frac{i_t^2}{2} + E_t U_g\left(G_{t+1}\right) + E_t \beta U_c\left(c_{t+1}, r_{t+1}, i_t\right) \quad (1) \\ s.t. c_{t+1} &= i_t\left(1 - \tau_{t+1}\right) \end{aligned}$$

U_g is the utility derived from public consumption while U_c is the utility derived from consumption and $\beta \in (0, 1]$ is the discount factor. In order to allow *loss aversion*, we let U_c depend not only on the level of private consumption but also on a reference point r_{t+1} and investments in the previous period.

The tax rate, τ_{t+1} is determined in the beginning of period τ_{t+1}, when investments i_t are sunk. Taxes are used for public good spending, having a marginal cost of unity to produce. The government budget constraint is therefore

$$G_t = i_{t-1}\tau_t$$

To simplify, we assume utility is linear in G_t with a marginal utility of $1+\gamma$, i.e.,

$$U_g(G_t) = (1+\gamma) G_t, \gamma > 0. \tag{2}$$

Note that γ reflects the rate of return of public investment. Given some value of marginal utility of private consumption, the relative value of public goods increase in γ.

2.1 Alternative interpretation

Above, we assumed that there is a single type of individuals in each cohort and that tax revenue is spent on a public good. An alternative interpretation of the model is that there are two types of individuals, entrepreneurs and workers. The former are like the ones described above while the latter are unable to undertake entrepreneurial activities (investments). Due to lower endowments than the entrepreneurs, workers have a higher marginal utility of consumption, namely $1+\gamma$,but have a reference point of consumption at zero. In the political process, to be described below, workers and entrepreneurs participate in voting on equal grounds.

2.2 Loss aversion

We assume that individuals have *loss aversion.* Following (Kahneman and Tversky 1991)), we note that the key ingredients of loss version are that

1. individuals care strictly more about losses relative to a certain reference point than about gains, i.e., there is first-order risk aversion around the reference point,

2. reference points are not fixed over time but is affected by individual experiences and expectations, and

3. individuals are risk loving in losses in the sense that

$$pU_c(r-x,r,i)+(1-p)U_c(r,r,i)>U_c(r-px,r,i)\,\forall x>0,p\in(0,1)\,. \tag{3}$$

According to (Quattrone and Tversky 1988), these two implications are key to understand the consequences of loss aversion for voting outcomes. Thus, our preferences must allow for these two features. However, since previous work (e.g., (Hassler, Rodríguez Mora, Storesletten and Zilibotti 2003)) has shown that (piecewise) linear utility makes it possible to analytical characterize Markov equilibria in dynamic political economy models, we want this here. We want to stress that loss-aversion is possible to model while restricting utility to be piece-wise linear. Specifically, we assume that

$$U_c(c_t,r_t,i_{t-1})=c_t-h\cdot I(c_t<r_t)\,i_{t-1} \tag{4}$$

Here, $I()$ is an indicator function that is unity if the argument is true and zero otherwise. As (Köszegi and Rabin 2006), we assume that utility is not only a function of the deviation of consumption from the reference point. Specifically, the first term in (4) represents the pure utility of consumption, while the second represents loss-aversion and is a function of consumption relative to the reference level. The parameter $h\geq 0$ measures the degree of loss-aversion and for $h=0$, utility is linear with a unitary marginal utility.

An important implication of our preference should be noted; According to (4), the utility loss associated with consumption falling below the reference point (in our case, taxes being too high), depends positively on the investment the individual has done. We label this feature *fairness*. In plain words, it implies that an individual that invested little (a lot) and gets fooled in the sense of getting to keep less of the return than implied by the reference point, will feel a smaller (larger) disutility or anger. In particular, our formulation implies that in the limit, as investments and consumption go to zero, the disutility of being taxed too heavily goes to zero.[3]

In Figure 1, we plot U_c against c, for $h, i > 0$ and a given value of r. We have also included a more "standard" continuous loss-averse utility function.

Clearly, our preference formulation induces *first-order* risk-aversion around the reference point. A better than fair bet, scaled by a factor k will always be *rejected* for sufficiently small k. Second, the preferences imply *risk-loving* behavior for losses – equation (3) is satisfied. Thus, the key implications of loss-aversion are also implications of our preference formulation.[4]

An individual with the preferences specified by (4) is *loss-averse* but risk neutral. We argue that loss-aversion and risk-aversion are quite different concepts and there seems to be no conceptual difficulty in allowing risk-neutral individuals to be loss-averse. A loss averse risk neutral individual do, however, certainly care about risks. In particular, since the individual has first-order risk aversion, she cares a lot about small risks. On the other hand, a mean preserving spread does not change expected utility as long

[3] We considered two alternatives to this assumption. First, one could assume that the loss is proportional to actual consumption c. But then, a tax rate of 100% removes the effects of loss aversion which is unreasonable. The second alternative we considered is to assume that the loss is a constant. As shown below, such an assumption complicates the analysis since it makes the political payoff function non-linear (but affine) in investments. However, we conjecture that such an assumption would produce similar results.

[4] In particular, assumption 1-4 in (Bowman, Minehart and Rabin 1999) are satisfied, except that we have replaced strong concavity by weak above the reference point and that we allowed loss-aversion to operate through a discontinuity at $c - r = 0$.

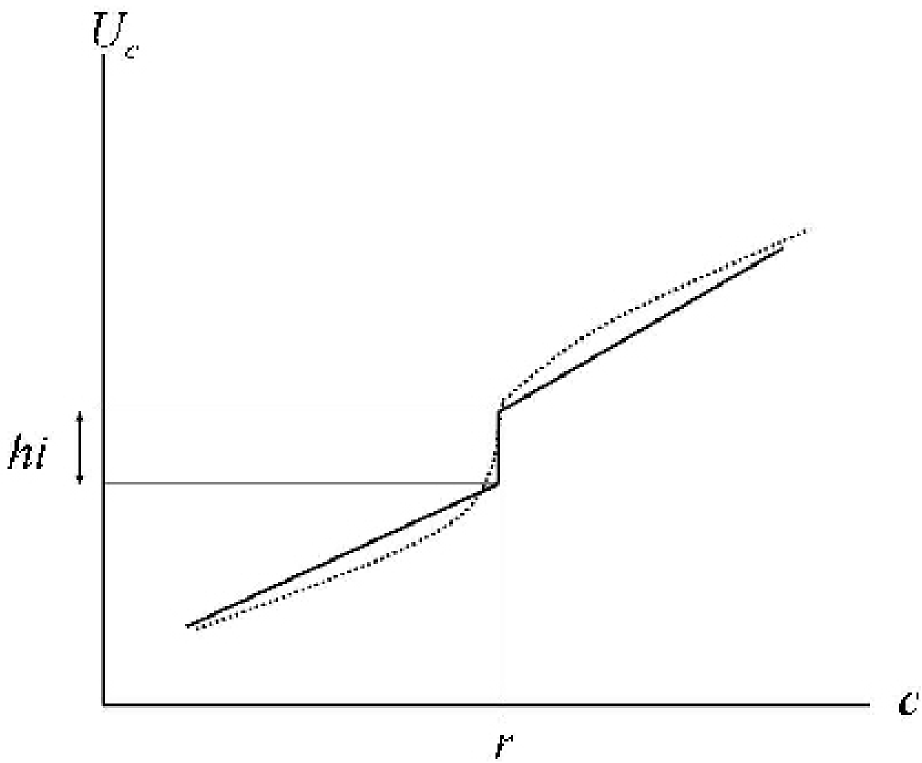

Figure 1: Figure 1. Risk-neutral, loss averse (solid) v.s. "standard" loss-averse utility (dotted).

as the probability of a loss is unchanged. That a mean preserving spread reduce expected utility is a key feature of risk-aversion – thus, we prefer to label these preferences as risk neutral.

Finally, let us comment on the choice of letting loss-aversion operate through a discontinuity at r. We do this for simplicity, believing that our results do not hinge on this assumption. Specifically; fix a small but strictly positive ε. It then seems safe to conjecture that all our results below would go through also if preferences were piecewise linear, given by the continuos function

$$U_c\left(c_t, r_t, i_{t-1}\right) = \begin{cases} c_t \text{ if } c_t \geq r_t, \\ c_t - h \cdot I\left(c_t < r_t\right) i_{t-1} \frac{r_t - c_t}{\varepsilon} \text{ if } c_t - r_t \in (-\varepsilon, 0), \\ c_t - h \cdot I\left(c_t < r_t\right) i_{t-1} \text{if } c_t - r_t \leq -\varepsilon. \end{cases} \tag{5}$$

2.3 Reference point dynamics

Our focus in this paper is on dynamic effects of loss-aversion. Unfortunately, the literature on prospect theory does not share this focus and much is therefore yet to be explored about what drives changes in the reference point over time. Bowman et al. (1999) construct a two-period model in which the first period reference point is exogenous while it in the second period is a weighted average of the first period's reference point and first period consumption. In this way, the author's can vary the degree of history dependence by changing the relative weights on the two determinants of the reference point.

In any case, it seems reasonable that the reference point should be positively affected by the individual's investment level. If an individual invests a lot, we believe that *ceteris paribus,* she expects to be able to consume more and perceives a loss if she is deprived of this. Specifically, we therefore assume that

$$r_{t+1} = i_t\left(1 - \tau_{t+1}^r\right), \tag{6}$$

where τ_{t+1}^r is a period t determined reference level for the period $t+1$ tax-rate. Before discussing how τ_{t+1}^r is determined, we want to stress that our assumption here makes the theory conceptually different from habit formation. In our model, the reference point is determined by the investment level. Under habit formation, if the habit and the investment level are directly related, the causal effect is rather in opposite direction, namely that the individual invests a lot because the habit for consumption is high.

In a dynamic political context there are reasons to believe the determination of τ^r_{t+1} is backward looking, but also there are reasons to consider the opposite, a forward looking reference point formation. The latter rests on the observation that political promises appear to have a large impact on outcomes and rhymes with the notion that individuals dislike being fooled in the sense of not given what they where promised. On the other hand, the evidence in (Quattrone and Tversky 1988) suggest a *status quo* bias consistent with backward looking reference point formation where precedence and tradition affects level of return an investor feels "entitled" too. We have no reason to discard any of these arguments and we will therefore consider both cases.

The first case is that the reference point is forward-looking and independent of the past. Furthermore, we assume that individuals have rational expectations implying that the reference point for $t+1$ must an equilibrium at $t+1$. The second, polar opposite case, is that the reference point is fully backward looking, namely that τ^r_{t+1} is determined by the current tax rate.

Clearly, forward-looking reference points may imply a multiplicity of equilibria. To shrink the set of equilibria, we assume that the political candidates, or the government, can make an announcement of its intentions for next period's tax rate. We denote the period t announcement of next period's tax rate τ^p_{t+1} and call it a *promise* although no exogenous commitment mechanism prevents the candidates from reneging on their promises. If the promise is an equilibrium, it is used to form a reference point for consumption. If the promise is not an equilibrium, we assume that individuals believe that next periods tax-rate is unity. An alternative interpretation of this is that the economy manage to achieve the *best* equilibrium, i.e., the reference point for next periods taxes is the one in the set of self-fulfilling equilibria that maximizes welfare. This interpretation is in line with (Köszegi and Rabin 2006) who focus on what we call forward looking reference points and defines preferred equilibrium as the one that maximize individual utility.

Our two assumptions are therefore;

Assumption F: Forward-looking reference points – $\tau_{t+1}^{r}=\tau_{t+1}^{p}$ if τ_{t+1}^{p} is an equilibrium. Otherwise, τ_{t+1}^{r} equals the highest equilibrium tax for $t+1$.

Assumption B: Backward-looking reference points – $\tau_{t+1}^{r}=\tau_{t}$.

3 Tax-determination

As in all politico-economic environment two sets of decisions are taken. Taxes are choose collectively and in a centralized manner via a certain political mechanism. The decisions here are taken strategically, internalizing aggregate and strategic effects. In addition each agent takes an individual investment decision in a decentralized manner, behaving atomistically taking aggregate current and future variables as given.

Collective decision – taxes. We assume that the political process is such that policies are chosen every period t in order to maximize the following function:

$$W_t \equiv U\left(c_t, r_t, i_{t-1}\right)+(1+\gamma)\,G_t-\frac{i_t^2}{2}+\beta E_t U\left(c_{t+1}, r_{t+1}, i_t\right)+(1+\gamma)\,G_{t+1},$$

subject to the resource constraints

$$\begin{aligned} G_t &= \tau_t i_{t-1}, \\ G_{t+1} &= \tau_{t+1} i_t, \\ c_t &= i_{t-1}\left(1-\tau_t\right), \\ c_{t+1} &= i_t\left(1-\tau_{t+1}\right). \end{aligned} \tag{7}$$

There are two interpretations of this. It can be read as the outcome of a political process characterized by probabilistic voting, with equal weights for all individuals and without the possibility of commitment. Alternatively it can be read as indicating that the taxes are set by a benevolent planner who maximizes average expected utility of living individuals without commitment.

Private decision – investment We require that investments are chosen rationally by the individuals, observing current tax rates and taking the expected tax-rate next period parametrically, i.e.,

$$i_t = \arg\max_{i_t} \left(-\frac{i_t^2}{2} + \beta E_t U \left(i_t \left(1 - \tau_{t+1}\right), r\left(i_t, \tau_{t+1}^r\right), i_t \right) \right). \tag{8}$$

3.1 Taxes under commitment

Let us begin by finding the value of taxes that maximize political welfare in steady state if there is commitment. The the political payoff is $i_{t-1}\left(1+\gamma\tau_t\right) - i_t^2 + \beta i_t \left(1+\gamma\tau_{t+1}\right)$ with $i_t = \beta\left(1-\tau_{t+1}\right)$ so the payoff is

$$i_{t-1}\left(1+\gamma\tau_t\right) - \frac{\left(\beta\left(1-\tau_{t+1}\right)\right)^2}{2} + \beta^2\left(1-\tau_{t+1}\right)\left(1+\gamma\tau_{t+1}\right)$$

This is maximized over τ_{t+1} by

$$\tau_c \equiv \frac{\gamma}{1+2\gamma}.$$

3.2 Equilibrium

We will consider both the finite and the infinite horizon case. In the former, we assume that no young is born in the final period T, so

$$W_T = U\left(c_T, r_T, i_{T-1}\right) + \left(1+\gamma\right) G_T \tag{9}$$

where

$$G_T = \tau_T i_{T-1}.$$

We solve for the equilibrium using backward induction, and the equilibrium is a sequence of functions $\tau_{T-s} = \tau_{T-s}(i_{T-s-1}, r_{T-s})$ and $i_{T-S} = i_{T-s}(i_{T-s-1}, r_{T-s})$ and, under assumption **F** $\tau^p_{T-s+1} = \tau^p_{T-s}(i_{T-s-1}, r_{T-s})$ satisfying

$$\begin{aligned}\tau(i_{T-s-1}, r_{T-s}) &= \arg \max_{\tau_{S-t}\in[0,1]} W_{T-t},\\ i(i_{T-s-1}, r_{T-s}) &= \arg\max_{i_t} \left(-\frac{i_t^2}{2} + \beta E_t U\left(i_t(1-\tau_{t+1}), r\left(i_t, \tau^r_{t+1}\right), i_t\right)\right),\end{aligned}$$

and

$$\tau^p(i_{T-s-1}, r_{T-s}) = \arg \max_{\tau^p_{S-t+1}\in[0,1]} W_{S-t}.$$

In the infinite horizon case, we will focus our attention on the case of Markov perfect equilibria, that is, we will retain the assumption from the finite horizon case that taxes, investments and promises are functions of the state variables i_{t-1} and r_t and in addition impose that these functions are independent of time. As we will see, this may require a broadening of the policy space to mixed strategies. It should be noted that under the assumption $r_t = i_{t-1}(1-\tau^r_t)$, we can use $\{i_{t-1}, \tau^r_t\}$ as an equivalent set of state variables, which turns out to be more convenient.

4 Equilibria without loss aversion

It is straightforward to verify the following proposition.

Proposition 1 *If $h = 0$, the only finite horizon equilibrium feature $i_t = 0$ and $\tau_t = 1$ for all t.*

Proof: If in the last period, $i_{T-1} > 0$, the unique value of τ_T maximizing W_T is $\tau_T = 1$. From the private investment rule, $i_{S-1} > 0$ is then contradicted. Continuing by backward induction, we find that, $i_t = 0\ \forall t$, which is supported by $\tau_t = 1\ \forall\ t$, (weakly) maximizing W_t in each period.

4.1 Non-Markovian equilibria without loss-aversion

In this paper, our main finding is that loss-aversion can provide a type of limited commitment. The standard alternative explanation for why economies might avoid bad equilibria like the one in proposition 1, is that trigger strategies could be used. In order to analyze whether loss aversion and trigger strategies are observationally equivalent it is of interest to look at a broader class of equilibria in the case when $h = 0$ and when the horizon is infinite. Although the set of equilibria of course is large, we can use the method of (Abreu, Pearce and Stacchetti 1990) to characterize the *best* equilibrium, i.e., the one the maximizes the welfare of the living individuals. We do this by first finding the *worst* equilibrium. Clearly, this is the finite horizon equilibrium with zero investments. Using this equilibrium as "punishment", we can sustain better equilibria than the one with zero investments.

Suppose $\hat{\tau}$ is a steady state equilibrium with an associated investment level $i = \beta(1-\hat{\tau})$ and a political payoff

$$\beta(1-\hat{\tau})(1+\gamma\hat{\tau}) - \frac{(\beta(1-\hat{\tau}))^2}{2} + \beta^2(1-\hat{\tau})(1+\gamma\hat{\tau}).$$

Now, the best deviation from this is to set $\tau_t = 1$, realizing that in period $t+1$ the finite horizon equilibrium will prevail as punishment for the deviation. The deviation yields a payoff

$$\beta(1-\hat{\tau})(1+\gamma).$$

The difference between the steady state payoff and the deviation is $\beta(1-\hat{\tau})\left((1+\beta)(1+\gamma\hat{\tau}) - \frac{\beta(1-\hat{\tau})}{2} - (1+\gamma)\right)$, which is positive for $\hat{\tau}$ close to unity. Therefore, tax rates smaller than unity can be sustained by the threat of reverting to the finite horizon equilibrium. However, τ_c might not necessarily be achievable. Specifically, the value of not deviating is negative if $\beta < \frac{2\gamma}{1+2\gamma}$. If this is the case, the best equilibrium implies setting

$$\tau = \hat{\tau} \equiv \frac{2\gamma - \beta}{\beta + 2\gamma (1+\beta)},$$

which is decreasing in β and approach unity as $\beta \to 0$.

We conclude:

Proposition 2 *If $h = 0$, the best infinite horizon equilibrium feature*

$$\tau_t = \tau_b = \max\left\{\frac{2\gamma - \beta}{\beta + 2\gamma (1+\beta)}, \frac{\gamma}{1+2\gamma}\right\} \forall t.$$

Corollary 3 *If $\beta < \frac{2\gamma}{1+2\gamma}$, $\tau_b < \tau_c$.*

In the next subsections, we consider equilibria under loss-aversion, first under assumption **F** and then **B**.

5 Equilibria with loss-aversion

5.1 Forward-looking reference points

Consider now case **F,** i.e., when the reference point is forward-looking and independent of the past. Take first the finite horizon case. First, we note that

$$c_t < r_t \Leftrightarrow \tau_t > \tau_t^r.$$

In the final period, using the resource constraints, we can therefore write the objective function as

$$\begin{aligned} W_T &= c_T - h \cdot I(c_T < r_T)\, i_{T-1} + i_{T-1}\tau_T (1+\gamma) \\ &= i_{T-1}\left(1 + \gamma\tau_T - h(\tau_t > \tau_t^r)\right). \end{aligned}$$

Clearly,

$$\tau_T = \arg\max_{\tau_T} W_T = \begin{cases} \tau_T^r \text{ if } \tau_T^r \geq \tau^* \equiv 1 - \frac{h}{\gamma}, \\ 1 \text{ else,} \end{cases}$$

independently of period $T-1$ investments. Consequently, any announced tax rate at or above τ^* is believed and will under assumption **F** form the reference point for consumption. Thus,

$$\tau_T^r = \begin{cases} \tau_T^p \text{ if } \tau_T^p \geq \tau^*, \\ 1 \text{ else.} \end{cases}$$

and investments are given by

$$i_{T-1} = \begin{cases} \beta\left(1-\tau_T^p\right) \text{ if } \tau_T^p \geq \tau^*, \\ 0 \text{ else.} \end{cases} = i_{T-1}\left(\tau_T^p\right)$$

The period $T-1$ payoff is maximized by

$$\begin{aligned} \tau^p &= \max\left\{\tau_c, \tau^*\right\} \\ \tau_{T-1} &= \begin{cases} \tau_{T-1}^p \text{ if } \tau_{T-1}^p \geq \tau^*, \\ 1 \text{ else.} \end{cases} \end{aligned}$$

Continuing by backward induction and noting that neither of the equilibrium functions depend on i_t, but only on τ_t^p, we derive the following proposition.

Proposition 4 *Under assumption **F**, there is a unique equilibrium in the finite horizon case. This equilibrium is also a Markov equilibrium in the infinite horizon and features*

$$\begin{aligned} \tau_t &= \tau\left(\tau_t^p\right) = \begin{cases} \tau_t^p \textit{ if } \tau_t^p \geq \tau^*, \\ 1 \textit{ else.} \end{cases} \\ \tau_{t+1}^p &= \tau^p\left(\tau_t^p\right) = \max\left\{\tau_c, \tau^*\right\}, \\ i_t &= i\left(\tau_t^p\right) = \begin{cases} \beta\left(1-\tau_t^p\right) \textit{ if } \tau_t^p \geq \tau^*, \\ 0 \textit{ else.} \end{cases} \end{aligned}$$

Starting from any i_0, τ_1^p, the equilibrium tax-rate is $\max\left\{\tau_c, \tau^*\right\} \forall t > 1$.

We should note here that the equilibrium under loss-aversion is independent of β, while this is not the case when there is no loss-aversion. In particular, the tax-rate in the best equilibrium is weakly negative in β and strictly negative for $\beta < \frac{2\gamma}{1+2\gamma}$. This implies that with some loss-aversion and a low enough discount factor, the economy can reach a better equilibrium even if is restricted to Markov strategies than without loss aversion but with no restriction on the strategy space. We believe that this provides a way of empirically distinguishing loss-aversion from reputation.

5.2 Backward-looking reference points

Case **B,** when the reference point is backward looking is slightly more complicated to analyze since it, for obvious reasons, tends go generate more dynamics. As we will see, however, the equilibrium is tractable for at least an interesting subset of the parameter range.

We first, we analyze the finite horizon equilibrium. Recall that in this case, reference taxes are backward-looking ($\tau_t^r = \tau_{t-1}$) and therefore, the choice of τ_t will generally depend on τ_{t-1}, which becomes the relevant state variable.

In the final period T;

$$\tau_T = T_T\left(\tau_{T-1}\right) = \begin{cases} \tau_{T-1} \text{ if } \tau_{T-1} \geq \tau^*, \\ 1 \text{ else,} \end{cases}$$

exactly as in the forward-looking case, except that τ_T^p is replaced by τ_{T-1}. Knowing this, individuals in period $T-1$ choose

$$i_{T-1} = i_{T-1}\left(\tau_{T-1}\right) = \begin{cases} \beta\left(1-\tau_{T-1}\right) \text{ if } \tau_{T-1} \geq \tau^*, \\ 0 \text{ else,} \end{cases}$$

where we note that $i'_{T-1}\left(\tau_{T-1}\right)$ is strictly negative in the range $\left[\tau^*, 1\right]$.

Consider now period $T-1$. Clearly, a motive to restrain current taxation has now arisen since by reducing τ_{S-1} from unity, current investments increase from 0 since $i'_{S-1}\left(\tau_{S-1}\right) < 0$. Specifically, while a tax-rate of unity would maximize the value of consumption of public and private goods in the current period, the welfare of the young able could be enhanced to the extent that restraining current taxes constrains next periods taxes. The important difference here compared to the analysis under forward-looking reference points is that now, there is an up-front cost of restricting next periods taxes since it means that the current stock of inelastic capital cannot be fully exploited.

In the somewhat uninteresting case, $\frac{h}{\gamma} > 1$, i.e., when full commitment can be achieved, either τ_{S-1} satisfies an interior first-order condition or is set equal to τ_{S-2} (see appendix for details, TBW). We will focus on a more interesting case, when loss aversion is not strong enough to completely "bind the hands" of the period T government. Specifically we in what follows make the assumption:

Assumption h $\frac{h}{\gamma} < \frac{\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)} \leq 1$.

In the appendix we show that ;

Lemma 5 *Under assumption* ***h****,* $T_{T-1}(T_{T-2}) = \tau^* \forall \tau_{S-2}$

The intuition for this result can be presented as follows. First, under assumption **h**, welfare falls in τ_{T-1} in the whole range $\tau_{T-1} \in [\tau^*, 1]$, provided i_{T-2} is done rationally, i.e., where a reduction in the current tax rate thus increases investment. In other words, h is not large enough to make it possible for the period $T-1$ government to achieve its most preferred tax, given that it could set the same tax in the current and next period. Thus, the equilibrium tax rate in period $T-1$ cannot be larger than τ^*.

To illustrate that the range of considered values of $\frac{h}{\gamma}$ is not trivially small, we plot $1 - \frac{\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}$, which is the lowest value of τ^* under assumption h, for two values of γ in figure 2.

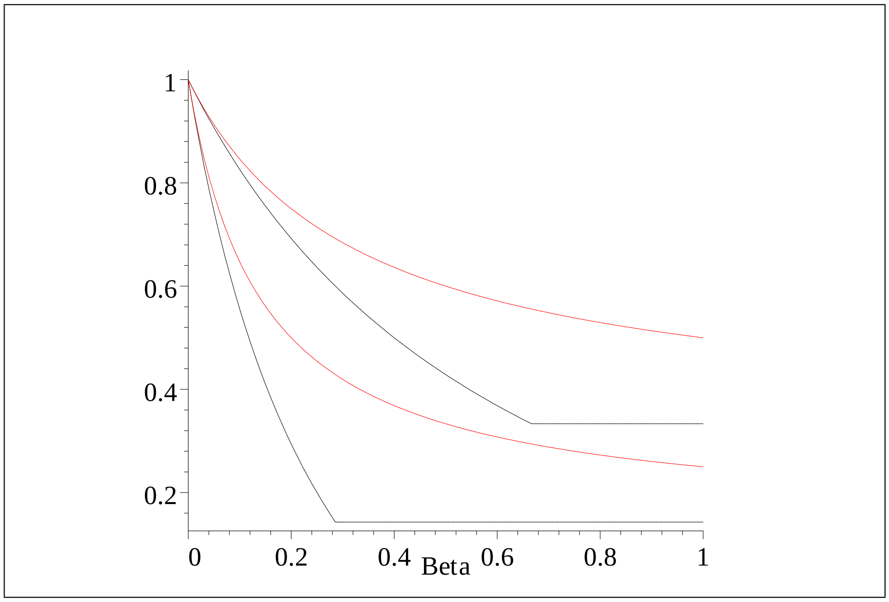

Fig 2. $1 - \frac{\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}$ for $\gamma = 0.2$ (bottom curve) and 1. Kinked curve represents best trigger strategy tax rate.

Second, under criterion assumption **h,** h is small enough for the government in period $T-1$ always to prefer to increase the tax rate to τ^*, if $\tau_{T-2} < \tau^*$, recognizing that this entails a cost due to loss-aversion and that if $\tau_{T-1} < \tau^*$, $I_{S-1} = 0$. Note that the fact that $i_{T-1}(\tau_{T-1})$ is discontinuous at $\tau_{T-1} = \tau^*$, is the reason for why it is optimal to increase taxes to τ^* also if τ_{T-2} is arbitrarily close to τ^*. If τ_{T-1} were to be set strictly below τ^*, individuals would know that in period T, the temptation to set $\tau_T = 1$, would not be resisted and therefore, $i_{T-1} = 0$, for all $\tau_{T-1} < \tau^*$. For this reason, the equilibrium policy in period $T-1$ cannot be to set $\tau_{T-1} < \tau^*$.

Since we have established that first equilibrium $\tau_{T-1} \leq \tau^*$ and second, equilibrium τ_{T-1} cannot be smaller than τ^*, the equilibrium policy in $T-1$ is clearly pinned down to $\tau_{T-1} = \tau^*$, independently of τ_{T-2}. In other words, the government in period $T-2$ *cannot* affect the period $T-1$ tax-rate. Therefore, the the problem of the period $T-2$ government is simply to maximize *current pay-off,* i.e., they face an identical problem as the period T government. Consequently, $T_T(\tau_{T-1})$ is optimal also in period $T-2$. By continuing this induction, we establish:

Proposition 6 *Under assumption **h**, the only finite horizon equilibrium features*

$$\begin{aligned}\tau_{T-s} &= T_{T-s}(\tau_{T-s-1}) \\ &= \begin{cases} T_e(\tau_{T-s-1}) = \begin{cases} 1 & \text{if } \tau_{T-s-1} < \tau^* \\ \tau_{T-s-1} & \text{if } \tau_{S-s-1} \geq \tau^* \end{cases} & \text{and } s \text{ is even.} \\ T_o(\tau_{T-s-1}) = \tau^* \quad \text{if} \quad s \text{ is odd.} \end{cases}\end{aligned}$$

This equilibrium involves an oscillation between forward-looking strategic behavior (the odd strategy) and complete "myopic" behavior, constrained by the previous tax rate. It is clear that these oscillations are key to the existence of the equilibrium. To see this, note that if a government (in period t) expects next government to behave strategically, by limiting τ_{t+1} in order to constrain later taxes, there is no need to be strategic already in period t. On the contrary, it is in this case superior to *procastinate* and make the myopically optimal decision today – the expectation of future strategic behavior, eliminates the need to be strategic today. Correspondingly, the expectation about future governments to behave myopically, creates an incentive to act strategically already in the current period. As we will discuss more below, we believe that this interaction between myopic and strategic behavior is necessary whenever commitment entails a short run-cost.

We should also note that although the tax policies must oscillate in equilibrium, the actual tax-rate does not. In fact, the tax-rate is constant at $1-\frac{h}{\gamma}$ after the first period.

5.3 Infinite horizon

First, we note extending the horizon backwards to infinity, the equilibrium described in proposition 6 does obviously not converge to a Markov-equilibrium in pure strategies. However, the logic behind the finite-horizon equilibrium – that expectation of future myopia breeds strategic behavior and *vice versa* –

suggests the existence of a Markov equilibrium in mixed strategies in an infinite horizon game *without a fixed endpoint.*. We may thus conjecture that if next periods government mixes between a myopic and a strategic policy with the right probabilities, we could make the current government indifferent between the same two policies. It turns out that this conjecture is correct and we can establish the following proposition.

Proposition 7 *Under assumption **h**, a Markov equilibrium exists with the following characteristics.*

$$\tau_t = T(\tau_{t-1}) = \begin{cases} T_e(\tau_{t-1}) & \text{with probability } 1 - p(\tau_{t-1}) \\ T_o(\tau_{t-1}) & \text{with probability } p(\tau_{t-1}) \end{cases}, \text{and}$$

$$i(\tau_t) = \begin{cases} 0 & \text{if} \quad \tau_t < \tau^* \\ \beta(1 - \tau_t + p(\tau_{t-1})(\tau_t - \tau^*)) & \text{if} \quad \tau_t \geq \tau^* \end{cases},$$

with $i'(\tau_t) < 0 \ \forall \tau_t > \tau^*$ *and where*

$$p(\tau) = \begin{cases} 1 \text{ for } \tau = 1. \\ \frac{p_1 + \frac{1}{2}\sqrt{\left((2p_1)^2 - 4p_2(2(p_1 + \gamma h) - p_2)\right)}}{p_2} \text{ for } 1 > \tau > \tau^* \\ < \gamma \text{ for } \tau < \tau^* \end{cases},$$

for

$$\begin{aligned} p_1 &= \gamma(\gamma(1+\beta)(1-\tau) - \beta\tau(1+\gamma) - h) \ \text{ and} \\ p_2 &= \beta(\gamma(1-\tau) - h)(1+2\gamma). \end{aligned}$$

Starting from any $\tau_0 \in [0,1]$ *and* $i_0 = i(\tau_0)$*, the equilibrium tax-rate converges with probability 1 to* τ^*.

Sketch of Proof: We begin by showing that if $\tau_{t-1} = \tau^*$,it is optimal to set $\tau_t = 1 - \frac{h}{\gamma}$, which is easy since both pure strategies prescribe this. Then, we note that if $\tau_{t-1} < \tau^*$, $i_{t-1} = 0$ so for the government to want to set $\tau_t = 1$, it has to be that this does not affect the future negatively, implying $p(1) = 1$. Finally, we solve for the function $p(\tau)$ such that for any $\tau_{t-1} \in (\tau^*, 1)$, the period t government is indifferent between the two pure strategies. Details in the appendix.

As an illustration, let us draw the pay-off to a period t government as a function of its choice of τ_t, given $\tau_{t-1} = 0.6$, when we set $h = .1, \gamma = .2, \beta = .75$. We show the value of the political objective function $W(\tau_t; \tau_{t-1}, i_{t-1})$ for all $\tau \in [0,1]$ in Figure 2.

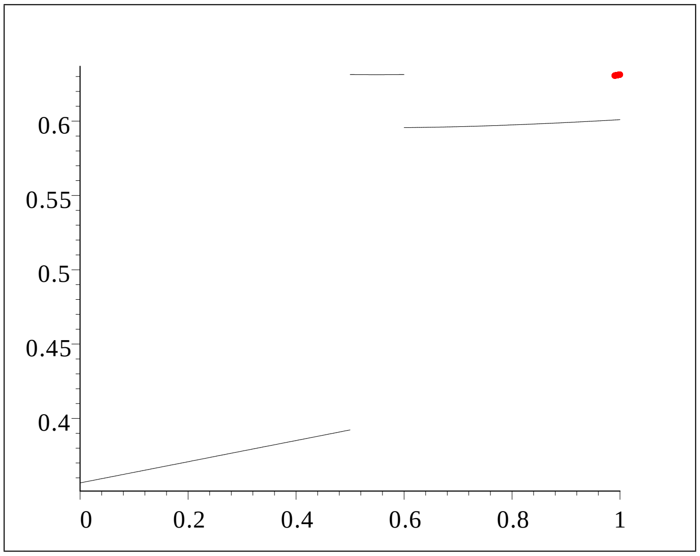

Figure 2. Political payoff against the choice of the current tax rate τ_t for $\tau_{t-1} = 0.6$..

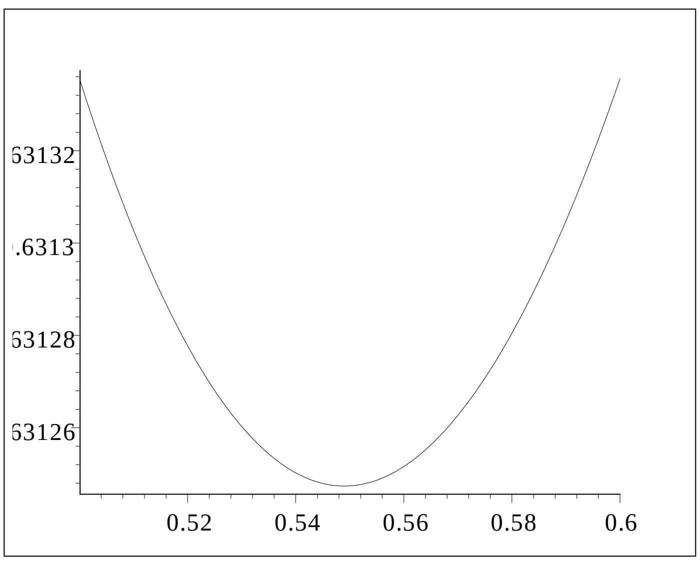

Figure 3. Political payoff against the choice of the current tax rate τ_t for $\tau_{t-1} = 0.6$.

While the function looks almost flat for the segment $\tau \in [\tau^*, \tau_{t-1}]$, a closer inspection, shown in Figure 3, illustrates that this is not the case. The government is indifferent between τ^* and τ_{t-1} and strictly prefers these values to any other $\tau \in [0, 1]$ .The $p(\tau)$ function is plotted in Figure 4.

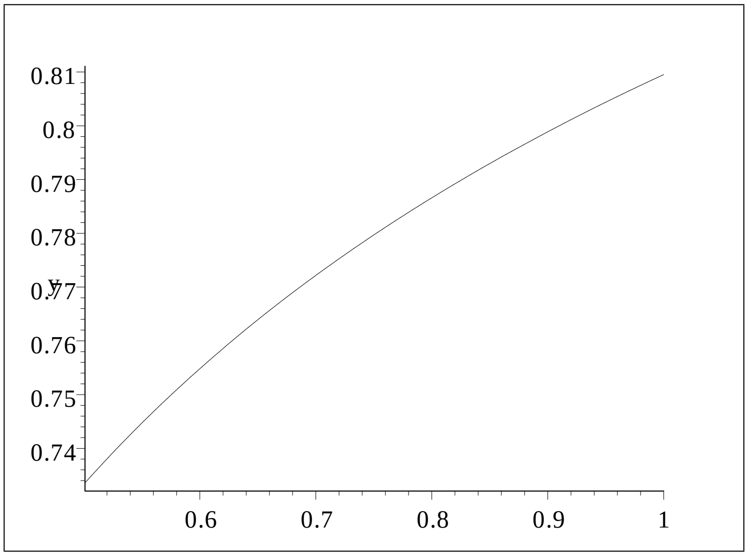

Figure 4. Mixing probability on τ^*.

6 Concluding remarks

We have in this paper provided a theory for how prospect theory can be used in political economy. In particular we want to emphasize the interpretation that people dislike being fooled and that this can help mitigate commitment problems that otherwise could have severe implications for society. Our model is simple and stylized in order to achieve analytical tractability. In particular when a new economic mechanism is analyzed, analytical transparence appears valuable. In future work, we plan to develop the model, including in particular, stochastic elements. We believe that the stochastic properties of the model can help distinguish this theory from alternative explanations for how commitment problems are overcome. With stochastics, promises will sometimes be broken and utility losses incurred. Under trigger strategies, such events may in some circumstances of asymmetric information lead to a switch to a more worse equilibrium. In our model, such a switch should not occur and history might not matter much at all. Therefore, the strong history dependence might be a way to distinguish the mechanisms, at least under forward looking reference point formation in which case the history is irrelevant.

Another issue is that we would like to analyze political competition involving an agency problem between voters and political executives. We have assumed away this by assuming probabilistic voting, which essentially means that tax rates are determined at the election date. In practice, political promises are, of course, most often seen in cases where a political candidate makes promises about what to do *after* being elected. We think similar mechanisms as the ones analyzed in this paper, may help politicians to make such promises credible.

We have in this paper assumed that loss-aversion applies only to private consumption, not to public. We believe that this can be defended on exactly the ground that private consumption is private. However, the empirical foundation for this presumption needs to be explored or alternatively, we may need to include loss aversion also in government spending. Furthermore, we believe that a possible way to analyze the how different issues can become salient in political campaigns is to assume that political parties are competing in affecting reference points in different political dimensions. Different parties may have different preferred dimensions, due to differences in ideology or competence, and may then prefer to establish reference points in different dimensions.

References

Abreu, D., Pearce, D. and Stacchetti, E.: 1990, Toward a theory of discounted repeated games with imperfect monitoring, *Econometrica* **58**(5), 1041–63.

Barro, R. J. and Gordon, D. B.: 1983, A positive theory of monetary policy in a natural rate model, *Journal of Political Economy* **12**(1), 101–21.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R.: 1997, A test of the theory of reference-dependent preferences, *Quarterly Journal of Economics* **112**(2), 479–505.

Bowman, D., Minehart, D. and Rabin, M.: 1999, Loss aversion in a consumption-savings model, *Journal of Economic Behavior and Organization* **38**, 155–178.

Cabrales, A., Nagel, R. and Mora, J. V. R.: 2006, It is hobbes, not rousseau: An experiment on social insurance. Mimeo: Universitat Pompeu Fabra.

Hassler, J., Rodríguez Mora, J. V., Storesletten, K. and Zilibotti, F.: 2003, The survival of the welfare state, *American Economic Review* **93**(1), 87–112.

Kahneman, D. and Tversky, A.: 1991, Loss aversion in riskless choice: A reference-dependent model, *Quarterly Journal of Economics* **106**, –.

Köszegi, B. and Rabin, M.: 2006, A model of reference dependent preferences, *Quarterly Journal of Economics* . Forthcoming.

Kydland, F. and Prescott, E.: 1977, Rules rather than discretion: The inconsistency of optimal plans, *Journal of Political Economy* **85**(3), 473–91.

Quattrone, G. A. and Tversky, A.: 1988, Contrasting rational and psychological analyses of political choice, *The American political science review* **82**(3), 719–736.

Rogoff, K. S.: 1985, The optimal degree of commitment to an intermediate monetary target, *Quarterly Journal of Economics* **100**(4), 1169–89.

7 Appendix

7.1 Proof of Lemma 5

The payoff in period $T-1$ is

$$\begin{aligned} W_{T-1} &= i_{T-2}\left(1+\gamma\tau_{T-1}-h\left(\tau_{T-1}>\tau_{T-2}\right)\right) \\ &\quad +\beta\left(i_{T-1}\left(\tau_{T-1}\right)\left(1+\gamma T_T\left(\tau_{T-1}\right)-h\left(T_T\left(\tau_{T-1}\right)>\tau_{T-1}\right)\right)\right)-\frac{i_{T-1}\left(\tau_{T-1}\right)^2}{2}. \end{aligned}$$

This is

$$\begin{aligned} & i_{T-2}\left(1+\gamma\tau_{T-1}-h\left(\tau_{T-1}>\tau_{T-2}\right)\right) \\ & +\begin{cases} \beta\left(\beta\left(1-\tau_{T-1}\right)\left(1+\gamma\tau_{T-1}\right)\right)-\frac{\left(\beta\left(1-\tau_{T-1}\right)\right)^2}{2} & \text{if } \tau_{T-1}\geq 1-\frac{h}{\gamma} \\ 0 \text{ else} \end{cases} \\ = \; & i_{T-2}\left(1+\gamma\tau_{T-1}-h\left(\tau_{T-1}>\tau_{T-2}\right)\right) \\ & +\begin{cases} \frac{\beta^2}{2}\left(1-\tau_{T-1}\right)\left(1+\left(1+2\gamma\right)\tau_{T-1}\right) & \text{if } \tau_{T-1}\geq 1-\frac{h}{\gamma} \\ 0 \text{ else} \end{cases} \end{aligned}$$

We will first show that under assumption **h**, there is no interior solution in the range $\tau_{T-1}\in\left[1-\frac{h}{\gamma},1\right]$ to the problem of maximizing W_{T-1} that can be an equilibrium under rational expectations. To show this, suppose on the contrary, that such an interior solution exists. This it satisfies the first order condition

$$\frac{d\left(i_{T-2}\left(1+\gamma\tau_{T-1}\right)+\frac{\beta^2}{2}\left(1-\tau_{T-1}\right)\left(1+\left(1+2\gamma\right)\tau_{T-1}\right)\right)}{d\tau_{T-1}}=0$$

$$\Rightarrow \tau_{T-1}=\gamma\frac{i_{T-2}+\beta^2}{\beta^2\left(1+2\gamma\right)}.$$

For this to be an equilibrium, we also need that investments in period $T-2$ are rational, i.e., that

$$i_{T-2}=\begin{cases} \beta\left(1-\tau_{T-1}\right) & \text{if } \tau_{T-2}\geq\tau_{T-1}, \\ \max\left\{\beta\left(1-\tau_{T-1}-h\right),0\right\} & \text{else.} \end{cases}$$

implying

$$\tau_{T-1} = \gamma \frac{\beta(1-\tau_{T-1}) + \beta^2}{\beta^2(1+2\gamma)}$$
$$\Rightarrow \quad \tau_{T-1} = 1 - \frac{\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}.$$

However, under assumption h, $1-\frac{h}{\gamma} > 1 - \frac{\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}$, so this interior equilibrium is not possible.

Alternatively, if $\tau_{T-2} < \frac{\gamma(1+\beta)}{\beta(1+\gamma)+\gamma(1+\beta)}$ and $\beta(1-\tau_{T-1}-h) \geq 0$,we have

$$\tau_{T-1} = \gamma \frac{\beta(1-\tau_{T-1}-h) + \beta^2}{\beta^2(1+2\gamma)}$$
$$\Rightarrow \quad \tau_{T-1} = 1 - \frac{\beta(1+\gamma)+\gamma h}{\beta + 2\beta\gamma + \gamma},$$

which also is below $1-\frac{h}{\gamma}$ under assumption **h**. Clearly, this is also the case if investments would have been zero in which case no temptation to set taxes above the first best $\frac{\gamma}{1+2\gamma} < 1 - \frac{h}{\gamma}$, exists.

We therefore conclude that there cannot be a rational expectations equilibrium in period $T-1$, where τ_{T-1} satisfies an interior first-order condition in the range $\tau_{T-1} \in \left[1-\frac{h}{\gamma}, 1\right]$.

Remaining possibilities, except our proposed equilibrium, where $\tau_{T-1} = 1 - \frac{h}{\gamma}$ for all τ_{T-2}, is that for some value of $\tau_{T-2} < 1 - \frac{h}{\gamma}$, it is not worth to take the cost due to loss aversion.

To see that this is not the case, we first note know that for τ_{T-2} in the range $[0, 1-\frac{h}{\gamma})$, the period T payoff is 0, if $\tau_{T-1}$is set equal to τ_{T-2} so payoff is $i_{T-2}(1+\gamma\tau_{T-1})$ which is increasing in τ_{T-1}. So the potential deviation from our equilibrium must be to set $\tau_{T-1} = \tau_{T-2}$. The payoff to this is $i_{T-2}(1+\gamma\tau_{T-2})$. Under the proposed equilibrium policy, $\tau_{T-1} = 1-\frac{h}{\gamma} > \tau_{T-2}$, so $i_{T-2} = \max\left\{\beta\left(\frac{h}{\gamma}-h\right), 0\right\}$, giving a deviation policy value no larger than $\beta\left(\frac{h}{\gamma}-h\right)(1+\gamma\tau_{T-2})$, which, of course, is increasing in τ_{T-1} so the supremum over all deviation policies is reached as τ_{T-2} approach $1-\frac{h}{\gamma}$, implying that the deviation payoff is bounded from above by $\beta\left(\frac{h}{\gamma}-h\right)\left(1+\gamma\left(1-\frac{h}{\gamma}\right)\right) \equiv W_{dev}$. We finally require that this is smaller than the payoff from the equilibrium policy of setting $\tau_{T-1} = 1 - \frac{h}{\gamma}$ for all τ_{T-2}. The payoff from this (when $\tau_{T-2} < 1 - \frac{h}{\gamma}$) is

$$i_{T-2}\left(1+\gamma\tau_{T-1}-h\right)+\frac{\beta^2}{2}\left(1-\tau_{T-1}\right)\left(1+\left(1+2\gamma\right)\tau_{T-1}\right)$$
$$\text{for } i_{T-2} = \beta\left(\frac{h}{\gamma}-h\right), \tau_{T-1}=1-\frac{h}{\gamma}$$

giving

$$\begin{aligned} & \beta\left(\frac{h}{\gamma}-h\right)\left(1+\gamma\left(1-\frac{h}{\gamma}\right)-h\right) \\ & +\frac{\beta^2}{2}\frac{h}{\gamma}\left(1+(1+2\gamma)\left(1-\frac{h}{\gamma}\right)\right) \\ \equiv\ & W_* \end{aligned}$$

and the condition for our proposed policy to be an equilibrium is thus

$$W_* - W_{dev} = -h\beta\left(\frac{h}{\gamma}-h\right)+\frac{\beta^2}{2}\frac{h}{\gamma}\left(1+(1+2\gamma)\left(1-\frac{h}{\gamma}\right)\right) \geq 0.$$

Setting the last LHS expression equal to zero gives a quadratic equation in h, with roots $h = 0$ and $h = \frac{2\beta\gamma(1+\gamma)}{2\gamma(1-\gamma)+\beta(1+2\gamma)} \equiv h_m > 0$. By differentiating $W_* - W_{dev}$ with respect to h at $h = 0$, we see that $W_* - W_{dev}$ is positive in the range $h \in [0, h_m]$. Finally, we need to establish that assumption **h** implies that $h \leq h_m$. To see this, we note that assumption **h** implies $h < \frac{\gamma\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)} \equiv h_h$, and we finally need to show that $h_m - h_h \geq 0$. At last,

$$h_m - h_h = \frac{\beta\gamma\left(1+\gamma\right)\left(\beta\left(1+2\gamma\right)+2\gamma^2\right)}{\left(2\gamma\left(1-\gamma\right)+\beta\left(1+2\gamma\right)\right)\left(\beta\left(1+\gamma\right)+\gamma\left(1+\beta\right)\right)}$$

For $\beta \in [0,1]$ the two real roots to $h_m - h_h = 0$ are $\gamma = 0$ and -1 and

$$\left[\frac{d\left(h_m-h_h\right)}{d\gamma}\right]_{\gamma=0} = 1.$$

Thus, $h_m - h_h > 0 \forall \gamma > 0$. QED.

7.2 Proof of proposition 7.

Noting that under the equilibrium policy,

$$E_t \tau_{t+1} = \begin{cases} \tau_t - p(\tau_t)(\tau_t - \tau^*) & \text{if } \tau_t \geq \tau^* \\ 1 - p(\tau_t)(1 - \tau^*) & \text{else} \end{cases}$$

Now, since $(p(1 - \tau^*) - h) < 0$, iff $p < \gamma$ which is the case under the equilibrium policy, equilibrium investments are

$$i_t(\tau_t) = \begin{cases} \beta(1 - \tau_t + p(\tau_t)(\tau_t - \tau^*)) & \text{if } \tau_t \geq \tau^* \\ 0 \text{ else.} \end{cases}$$

We define the political payoff by choosing τ_t, given τ_{t-1} and the equilibrium strategy is played in the future as

$$\begin{aligned} W(\tau_{t-1}, \tau_t) &\equiv i(\tau_{t-1})(1 + \gamma\tau_t - h(\tau_t > \tau_{t-1})) - \frac{i(\tau_t)^2}{2} \\ &\quad + \beta(i(\tau_t)(1 + \gamma E_t T(\tau_t) - hE_t(T(\tau_t) > \tau_t))) \end{aligned}$$

Let us now go over the value function in the different regions of τ_{t-1}. The proof will proceed by verifying that the equilibrium policy is optimal for all τ_{t-1}.Suppose first that $\tau_{t-1} < \tau^*$, then $i_{t-1} = 0$, and the equilibrium policy prescribes mixing between $\tau_t = \tau^*$ and $\tau_t = 1$. Clearly these choices are both optimal provided they both lead to $\tau_{t+1} = \tau^*$ which they do under the equilibrium policy.

Consider then the range $\tau_{t-1} \geq \tau^*$. Here, the equilibrium prescribes mixing between $\tau_t = \tau_{t-1}$ and $\tau_t = \tau^*$. Therefore, we need

$$W(\tau_{t-1}, \tau_{t-1}) = W(\tau_{t-1}, \tau^*)$$

Using $i(\tau) = \beta(1 - \tau + p(\tau)(\tau - \tau^*))$ and the definition of τ^* and to simplify notation using $\tau = \tau_{t-1}$

this yields the following second degree equation in p;

$$\begin{aligned} 0 &= p^2 + \frac{2}{1+2\gamma}\left(\frac{\gamma(1-\tau)-\tau(1+\gamma)}{\tau-\tau_s} - \frac{\gamma}{\beta}\right)p \\ &\quad - \frac{1}{1+2\gamma}\left(\frac{2(\gamma(1-\tau)-\beta(\tau(1+\gamma)-h))}{(\tau-\tau_s)\beta}+1\right). \end{aligned} \tag{10}$$

The relevant root is given by

$$p = p(\tau),$$

as defined in the proof.

In the range $\tau_{t-1} \geq \tau^*$, it now remains to be shown;

1. That $p(\tau) \in [0,1]$,
2. that no choice of τ_t below τ^* is optimal,
3. that no choice of τ_t above τ_{t-1} is optimal and
4. that no choice of τ_t in the range (τ^*, τ_{t-1}) is optimal.

1. To show that $p(\tau) \in [0,1]$ for all $\tau > \tau^*$ (remember that $p(\tau^*)$ is not defined and $p(\tau)$ for $\tau < \tau^*$ is only required to be smaller than γ), we first note that

$$\frac{p_1}{p_2} - \frac{1}{2}\sqrt{\left(2\frac{p_1}{p_2}\right)^2 - 4\left(2\frac{p_1+\gamma h}{p_2}-1\right)} > \frac{p_1}{p_2} - \frac{1}{2}\sqrt{\left(2\frac{p_1}{p_2}\right)^2} = 0,$$

since$\frac{d\left(2\frac{p_1+\gamma h}{p_2}-1\right)}{dh} < 0 \forall \tau_{t-1} > \tau^*$ and using assumption **h**, we have $\left[2\frac{p_1+\gamma h}{p_2}-1\right]_{h=\frac{\gamma\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}} = 1 + \frac{2\gamma}{(1+2\gamma)\beta} > 0$. Thus, $p(\tau) > 0$ for all $\tau > \tau^*$.

Second, $p(\tau)$ is smaller than unity if $\frac{p_1}{p_2} < 1$, i.e. if $p_2 - p_1 > 0$. Now, since $p_2 - p_1$ is increasing in τ and decreasing in h in the relevant range ($\tau > \tau^*$ and $h < \frac{\gamma\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}$) we have

$$p_2 - p_1 > [p_2 - p_1]_{\tau=1-\frac{h}{\gamma}, h=\frac{\gamma\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}} = \frac{\gamma^2\beta(1+\gamma)}{\beta(1+2\gamma)+\gamma} > 0.$$

2. This is immediate. Setting $\tau_t < \tau^*$ yields zero investment and lower current payoff $i_{t-1}(1+\gamma\tau_t)$ than τ^*.

Before going to part 3 and 4, we establish the following lemma.

Lemma 2. *Under assumption* ***h****,* $i'(\tau) < 0 \ \forall\tau \in (\tau^*, 1)$

Proof below.

We can now continue to point 3. Since we consider $\tau_{t-1} \geq \tau^*$, current payoff is $i_{t-1}(1+\gamma\tau_t - h(\tau_t > \tau_{t-1}))$ not higher by setting $1 > \tau_t > \tau_{t-1}$. Furthermore, any $\tau_t \in (\tau_{t-1}, 1)$ yields lower continuation payoff and must be suboptimal under lemma 1. Only setting $\tau_t = 1$, remains. Define the continuation payoff including current investments if future tax-rates are τ^* as

$$V^* \equiv -\frac{\left(\beta\frac{h}{\gamma}\right)^2}{2} + \beta\left(\beta\frac{h}{\gamma}(1+\gamma\tau^*)\right).$$

By setting $\tau_t = 1$, we get $i(\tau_{t-1})(1+\gamma-h)+V^*$. By following the equilibrium policy, e.g., by setting $\tau_t = \tau^*$, the payoff is $i(\tau_{t-1})(1+\gamma\tau^*)+V^*$ and by the definition of τ^* these payoffs are identical, so there is no gain to be made to deviate from the equilibrium by setting $\tau_t = 1$.

4. We have chosen p so that

$$\begin{aligned} i(\tau_{t-1})\gamma(\tau-\tau^*) &= -\frac{i(\tau^*)^2}{2} + \beta(i(\tau^*)(1+\gamma\tau^*)) \\ &\quad - \left(-\frac{i(\tau)^2}{2} + \beta(i(\tau)(1+\gamma(p(\tau)\tau^* + (1-p(\tau))\tau)))\right) \end{aligned}$$

if $\tau = \tau_{t-1}$.I.e., the short run temptation to set high taxes $(\tau_t = \tau_{t-1})$ is balanced by the long run gain of setting $\tau_t = \tau^*$. Now, given τ_{t-1}, could there be another $\tau \in (\tau^*, \tau_{t-1})$ that satisfies this? Suppose there is such a solution, and call it $\hat{\tau}$,then

$$\begin{aligned} i\left(\tau_{t-1}\right)\gamma\left(\hat{\tau}-\tau^{*}\right) &= -\frac{i\left(\tau^{*}\right)^{2}}{2}+\beta\left(i\left(\tau^{*}\right)\left(1+\gamma\tau^{*}\right)\right) \\ &\quad -\left(-\frac{i\left(\hat{\tau}\right)^{2}}{2}+\beta\left(i\left(\hat{\tau}\right)\left(1+\gamma\left(p\left(\hat{\tau}\right)\tau^{*}+\left(1-p\left(\hat{\tau}\right)\right)\hat{\tau}\right)\right)\right)\right) \end{aligned}$$

From the construction of p we know that

$$\begin{aligned} i\left(\hat{\tau}\right)\gamma\left(\hat{\tau}-\tau^{*}\right) &= -\frac{i\left(\tau^{*}\right)^{2}}{2}+\beta\left(i\left(\tau^{*}\right)\left(1+\gamma\tau^{*}\right)\right) \\ &\quad -\left(-\frac{i\left(\hat{\tau}\right)^{2}}{2}+\beta\left(i\left(\hat{\tau}\right)\left(1+\gamma\left(p\left(\hat{\tau}\right)\tau^{*}+\left(1-p\left(\hat{\tau}\right)\right)\hat{\tau}\right)\right)\right)\right) \end{aligned}$$

thus, we must have

$$i\left(\tau_{t-1}\right)=i\left(\hat{\tau}\right)$$

which is a contradiction since $i'\left(\tau\right)<0$ in the relevant range under lemma 1.

7.3 Proof of lemma 2.

Totally differentiating (10) yields

$$\frac{dp}{d\tau}=\frac{\left(1-p\right)\left(\tau_{s}+\gamma-2h\right)-\frac{h}{\beta}}{\left(\tau-\tau_{s}\right)^{2}\left(1+2\gamma\right)\left(p+\frac{1}{1+2\gamma}\left(\frac{\gamma(1-\tau)-\tau(1+\gamma)}{\tau-\tau_{s}}-\frac{\gamma}{\beta}\right)\right)}.$$

Therefore,

$$\begin{aligned} \frac{i'\left(\tau\right)}{\beta} &= -\left(1-p\right)+\left(\tau-\tau_{*}\right)\frac{dp}{d\tau} \\ &= -\left(1-p\right)\left(1-\frac{\left(\tau_{s}+\gamma-2h\right)-\frac{h}{\beta}}{\left(\tau-\tau_{s}\right)\left(1+2\gamma\right)\left(p+\frac{1}{1+2\gamma}\left(\frac{\gamma(1-\tau)-\tau(1+\gamma)}{\tau-\tau_{s}}-\frac{\gamma}{\beta}\right)\right)}\right) \\ &\equiv -\left(1-p\right)X \end{aligned}$$

Now, since X is increasing in p, we have

$$\begin{aligned} X &> \left(1 - \frac{\tau_s + \gamma - 2h - \frac{h}{\beta}}{(\tau - \tau_s)(1 + 2\gamma)\left(\frac{1}{1+2\gamma}\left(\frac{\gamma(1-\tau)-\tau(1+\gamma)}{\tau-\tau_s} - \frac{\gamma}{\beta}\right)\right)}\right) \\ &= \left(1 - \frac{\tau_s + \gamma - 2h - \frac{h}{\beta}}{\gamma - 2\gamma\tau - \tau - (\tau - \tau_s)\frac{\gamma}{\beta}}\right). \end{aligned}$$

The final expression is decreasing in τ, so

$$X > 1 - \frac{\tau_s + \gamma - 2h - \frac{h}{\beta}}{\gamma - 2\gamma - 1 - (1 - \tau_s)\frac{\gamma}{\beta}},$$

where the RHS is decreasing in h. Therefore,

$$X > \left[1 - \frac{\tau_s + \gamma - 2h - \frac{h}{\beta}}{\gamma - 2\gamma - 1 - (1 - \tau_s)\frac{\gamma}{\beta}}\right]_{h=\frac{\gamma\beta(1+\gamma)}{\beta(1+\gamma)+\gamma(1+\beta)}} = 1.$$

Consequently, $i'(\tau) < -\beta(1-p) < 0$.