Abstract

This paper develops a general theory for the design of retirement policies, like social security and retirement accounts, within a Mirrlees taxation framework with hidden present bias and sophistication. The paper shows how policies can utilize the time inconsistency of agents to improve welfare above the constrained efficient optimum. In particular, in an environment with both time-consistent and time-inconsistent agents, welfare increases monotonically with the population of time-inconsistent agents. For implementation, the paper focuses on the design of social security and retirement accounts. The optimal policy has social security benefits decreasing in progressivity with the initial withdrawal age. It also allows early withdrawals from retirement accounts only when there are large income discrepancies. The coexistence of both policies screens sophistication and present bias. These proposals outperform traditional policies, like linear savings subsidies or mandatory savings, by increasing output efficiency. (JEL Codes: D03, D62, D82, D84, D86, D91, H21)

Keywords: Retirement policy, Time inconsistency, Optimal taxation
1 Introduction

Policymakers and researchers have been concerned with the issue of inadequate retirement savings. In response, strengthening social security and increasing participation in retirement accounts are mentioned as core issues by the previous US administration. On the other hand, policymakers are also concerned about social insurance and the sustainability of these programs. The top two recommendations of the National Commission on Fiscal Responsibility and Reform (2010) are to make social security benefits more progressive with income, and to enhance the minimum benefits for low-wage workers. It also recommends increasing the maximum amount of taxable income for social security. For many, this reform will raise taxes and cut their social security benefits, which could introduce additional distortions to the labor supply. As a result, the trade-off between increasing retirement welfare and minimizing the cost of its provision is an urgent issue.

To study this question, this paper extends the Mirrlees taxation framework to incorporate quasi-hyperbolic discounting with heterogeneous present bias and sophistication. In essence, agents have private information about their productivity, the degree of present bias, and their beliefs regarding the severity of the bias. This paper provides a theoretical framework to designing retirement policies, which sheds light on features in social security and retirement accounts that could improve welfare. The key is in utilizing the agent’s time inconsistency, so policies go beyond mitigating the present bias. Traditional policies, such as linear savings subsidies or mandatory savings, increase savings by offsetting the bias independent of the asymmetric information. Using traditional policies, the government is able to guarantee the constrained efficient optimum, but cannot do better. I consider policy instruments that could elicit private information at a lower cost while mitigating the present bias. I also identify sufficient conditions such that the efficient allocation is implementable despite the presence of private information.

---

1The National Research Council (2012) finds that up to 2/3 of the US population is saving inadequately for retirement. Scholz, Seshadri, and Khitatrakun (2006), one of the more conservative studies, estimates at least 20% of the US population are not saving enough for retirement.

2See President Obama’s core issues (https://www.whitehouse.gov/issues/seniors-and-social-security). It also aims to use lessons from behavioral economics to attain its goals (Executive Order No. 13707, 2015).

3Auerbach et al. (2016) show the current fiscal system may encourage the elderly to retire early. It demonstrates how the optimal retirement policy needs to be considered in tandem with the design of the tax system.

4Empirical evidence shows that models with time-inconsistent preference can explain the consumption and savings patterns observed in the data (Angeletos et al., 2001; Laibson, Repetto, and Tobacman, 2007).

5By retirement policies, I am referring to social security, policies regarding retirement accounts, and any other programs related to retirement welfare.

6Efficient allocation satisfies consumption smoothing, full insurance and efficient production (See Proposition).
With time-inconsistent agents, the design of off-equilibrium path policies is important (Esteban and Miyagawa, 2005; Eliaz and Spiegler, 2006). It exploits the disagreement between present and future-selves. The contributions of this paper are twofold. First, this paper extends these methods to a Mirrlees taxation model, where agents have hidden present bias and sophistication. The gain in welfare from implementing off-equilibrium path policies in a Mirrlees setting is not straightforward. I show that if the utility from consumption is unbounded above and below, then the efficient allocation is implementable because the government is able to make arbitrarily large off-equilibrium threats and promises. I also argue how welfare can still be improved even when threats and promises are limited (see Section 7). Secondly, this paper also demonstrates how to implement off-equilibrium path policies in public finance and explores its implications on the design of retirement policies.

To see how off-equilibrium path policies can improve welfare, consider a three period model with two productivity types, productive and unproductive, and all agents in the economy share the same degree of present bias. Agents work in the first two periods and retire in the last period.

For fully na¨ıve agents (completely unaware of present bias), the government designs policies that improve output efficiency by promising retirement benefits that cover a portion of the information rent. If an agent produces a high output in the first period, then an option of claiming higher retirement benefits would be available in the second period. However, in the second period, the present-biased agent prefers immediate gratification over delayed benefits, so the agent would forego the option of claiming the higher retirement benefits. Fully naïve agents do not foresee this and choose to produce efficiently due to over-estimating the value of retirement benefits to their future-selves. I will refer to this mechanism as a betting mechanism, which is an extension of speculative bets in Eliaz and Spiegler (2006) and Heidhues and Koszegi (2010) to a public policy setting.

For sophisticated agents (fully aware of their bias and thus demand commitment), the government provides commitment in exchange for an increase in output efficiency. If an agent produces a low output in the first period, then the agent would face a menu of policies in the next period. One of the policies would exacerbate the present bias by increasing consumption in the second period at the expense of retirement welfare. A present-biased agent in the second period would prefer this policy from the menu. However, to qualify for it, the agent has to produce a sufficiently high output in the second period. Hence, a productive agent would face a policy that exacerbates the present bias if production was inefficiently low in the first period, while the unproductive agent would not find this policy appealing due to the required increase in output. By backward induction, a productive agent

---

7In a Mirrlees taxation model, the relevant deviation is for agents to mimic less productive agents.
is willing to exchange the information rent for commitment. I will refer to this mechanism as a conditional commitment mechanism, which is an extension of Esteban and Miyagawa (2005), Galperti (2015) and Bond and Sigurdsson (2017) to the Mirrlees environment. Both betting and conditional commitment mechanisms help screen productivity for partially naïve agents (aware of present bias but underestimate its severity), because they are willing to exchange information rents for either commitment or speculative bets.

The first main result shows that in a Mirrlees economy with time-inconsistent agents and hidden present bias and sophistication, off-equilibrium path policies can improve welfare above the constrained efficient optimum. The government can choose a target level of sophistication. It designs a betting mechanism for the least present-biased agents in the economy for the targeted sophistication. It also designs a conditional commitment mechanism for the most present-biased agents in the economy for the same targeted sophistication. The paper shows that agents more naïve than the targeted level would self-select into the betting mechanism, while agents more sophisticated would self-select into the conditional commitment mechanism. The betting mechanism is designed for the least present-biased agent, because off-equilibrium path policies have the information rents loaded in the future, so if the least present-biased agents prefer the on-equilibrium path policy, then more present-biased agents would prefer it too. On the other hand, the conditional commitment mechanism is designed for the most present-biased agent, because off-equilibrium path policies decrease retirement consumption, so if the most present-biased agents prefer the on-equilibrium path policy, then less present-biased agents would prefer it too. As a result, screening sophistication and present bias is costless. Under the assumption of unbounded utility, screening productivity is also costless, so the efficient allocation is implementable.

The second main result examines an economy with time-consistent agents. The presence of time-consistent agents limits the effectiveness of off-equilibrium path policies, because they do not require commitment and are not susceptible to speculative bets. Nevertheless, off-equilibrium path policies can still improve welfare by separating time-inconsistent agents from time-consistent agents and induce the time-inconsistent agents to increase output efficiency. This paper shows that welfare increases with the population of time-inconsistent agents. This is because total output increases with the proportion of present-biased agents, so the government can provide more information rent per time-consistent agent without additional distortions. This setting is important because off-equilibrium path policies may encourage agents to adopt outside commitment, and time-consistent agents could be in-

---

8This result differs from Eliaz and Spiegler (2006), where each specific sophistication level is provided with a different allocation. This is because their paper is exploitative, while the objective of this paper is paternalistic, so the government attempts to implement the same consumption for all agents regardless of sophistication or degree of present bias.
terpreted as time-inconsistent agents with commitment. I also show how the presence of time-consistent agents places a natural restriction on speculative bets, so conditional commitment mechanisms generally outperform betting mechanisms.

This paper provides new and intuitive insights on the design of policies without recommending significant qualitative changes to existing policies. It has been argued that people in the US are claiming social security benefits too early in life, and should instead retire later and delay benefits claiming. National Commission on Fiscal Responsibility and Reform (2010) recommends the use of behavioral economics, more specifically choice architecture, to nudge people to retire and claim benefits later. Contrary to this perspective, this paper suggests that social security benefits should decrease in progressivity with the initial age of claiming benefits. Non-sophisticated agents (fully naïve and partially naïve agents) would plan to claim at a later age and the decrease in progressivity would encourage them to work efficiently, which helps the sustainability of the social security system by increasing taxable income. However, they would claim earlier than planned and the more progressive benefits for early claimants improve insurance.

Sophisticated agents are concerned that they would withdraw early from their retirement accounts due to present bias, and therefore prefer illiquid accounts (Beshears et al., 2015a,b). I show that increasing the liquidity of retirement accounts to allow for early withdrawals helps implement the efficient allocation. The idea is to allow for early withdrawals only if the agent’s present income is significantly higher than past income. Agents who work efficiently would face illiquid retirement accounts in the future. Agents who work inefficiently would face a liquid account that tempts them to work more to withdraw early in the future, resulting in low savings for retirement. Thus, agents work efficiently for commitment. The paper shows how the redesigned retirement accounts and social security system can coexist to increase welfare through the costless screening of present bias and sophistication.

When time-consistent agents are present, I show how participation in retirement savings accounts should be voluntary. Time-inconsistent agents are encouraged to enroll in retirement accounts that provide them with commitment, but time-consistent agents should have full discretion on their savings. The enrollment decision of agents helps the government separate time-inconsistent agents from time-consistent agents. Income taxes depend

---


10Benefits are progressive if the ratio of lifetime benefits to lifetime payroll taxes is higher for low income individuals than it is for people with higher average income. Benefits decrease in progressivity if the difference of this ratio between the low income individuals and higher income individuals also decreases.

11This implementation is similar to the idea of partially illiquid assets in Bond and Sigurdsson (2017). However, they analyze an endowment model so the implementation is restricted to savings.

12This is consistent with current proposals by many US states, like Oregon and Illinois, to provide all employees with retirement savings accounts which they can choose to opt out.
on enrollment choice to implement the constrained efficient optimum.

1.1 Related Literature

There have been several papers examining the design of policies with behavioral agents. Farhi and Gabaix (2015) study optimal taxation (Ramsey, Pigou and Mirrlees) with behavioral agents in a static environment by using sparse maximization (Gabaix, 2014). They are able to derive general results without specifying the bias, so it could potentially be applied to environments with agents who suffer from a wide array of behavioral biases. Lockwood (2016) extends Farhi and Gabaix (2015) to present biased agents, and finds that the optimal marginal tax rate could be negative. Lockwood (2016) does not model dynamic consumption so the implementation focuses on income taxation, while this paper emphasizes consumption and savings with present bias and focuses on optimal retirement policies.

Guo and Krause (2015) study a dynamic Mirrlees environment with sophisticated time-inconsistent agents where the government does not have full commitment. Moser and de Souza e Silva (2017) consider a two-dimensional screening setup with hidden present bias and productivity and decentralizes the optimum using social security and retirement accounts. In Moser and de Souza e Silva (2017), present bias is stochastic, so providing a flexible retirement plan for high income agents while limiting the options for low income agents can relax incentive compatibility constraints and is optimal. This paper focuses on deterministic present bias, so flexible policies may induce under-savings, which is a distortion that could be avoided when agents have constant present bias.

In other related work, Diamond and Spinnewijn (2011) discuss a model with heterogeneity in both productivity and time preference (agents are time-consistent). Krusell, Kuruscu, and Smith (2010) study the optimal taxation of consumers who suffer from temptation in a complete information environment. Amador, Werning, and Angeletos (2006) examine government policies for agents who suffer from temptation and are subject to future taste shocks. Halac and Yared (2014) apply a repeated model of Amador, Werning, and Angeletos (2006) with persistent shocks. Similar to this paper, Halac and Yared (2014) also point out how allowing the government to revise past reports in the future can help relax incentive compatibility and deter misreporting in the present.

This paper is also related to several behavioral contracting papers. In particular, Esteban and Miyagawa (2005) examine optimal pricing schemes with time-inconsistent agents and find that distortions from information asymmetry can be averted when agents are tempted

---

13 Meier and Sprenger (2015) find evidence of present bias and obtain estimates for the quasi-hyperbolic model. They find the parameter estimates of $\beta$ and $\delta$ to be relatively stable over time, which supports the theoretical environment of this paper.
to over-consume. Bond and Sigurdsson (2017) demonstrate how off-equilibrium path options in commitment contracts can help time-inconsistent agents follow through with an ex-ante plan that accommodates their flexible needs. Eliaz and Spiegler (2006) examine a model with diversely naïve agents and found that firms can screen beliefs by bisecting the population into relatively sophisticated and relatively naïve agents. Similar to this paper, they find relatively sophisticated agents exert no informational externality on the relatively naïve agents. Galperti (2015) extends Amador, Werning, and Angeletos (2006) to a sequential screening model where a mechanism designer first screens time consistency and then the taste shock. In this paper, the government screens the agent’s private information (productivity, time-inconsistency and sophistication) simultaneously.

The paper is organized as follows. Section 2 presents the model. Section 3 presents the general mechanism. Section 4 examines the setting with heterogeneous present bias and sophistication. Section 5 considers the effects of time-consistent agents. Section 6 discusses a reform of the social security and retirement accounts. Section 7 discusses some extensions and impediments to the mechanism and Section 8 concludes.

2 The Model

A continuum of agents of measure one live for three periods: \( t \in \{0, 1, 2\} \). Agents are heterogeneous in productivity, which is denoted by \( \Theta = \{\theta_1, \theta_2, \ldots, \theta_M\} \), with \( \theta_{m+1} > \theta_m \) and \( |M| \geq 2 \). Productivity is distributed according to \( \Pr(\theta = \theta_m) = \pi_m > 0 \), for all \( \theta_m \in \Theta \) with \( \sum_{m=1}^M \pi_m = 1 \).

The production technology is linear and depends on labor input \( l_t \) and the productivity of the agent: \( y_t = \theta l_t \). Agents have access to a storage technology that transfers one unit of good in period \( t \) to one unit of period \( t + 1 \) good. The government does not observe \( \theta \) and \( l_t \), but it observes \( y_t \).

The period utilities \( u_t : \mathbb{R}_+ \mapsto \mathbb{R} \) are twice differentiable and \( u'_t, -u''_t > 0 \). The dis-utility from labor \( h_t : \mathbb{R}_+ \mapsto \mathbb{R} \) satisfies \( h'_t, h''_t > 0 \). Single crossing is automatically satisfied. For most of the paper, I will assume that \( u_t \) is unbounded below and above.

**Assumption 1** For any \( t \in \{1, 2\} \), \( u_t \) has full range (\( u_t (\mathbb{R}_+) = \mathbb{R} \)).

In Section 4 I show that Assumption 1 is a sufficient condition for the efficient allocation to be incentive compatible. In Section 7 I discuss how the efficient allocation would not be implementable when Assumption 1 fails, but the main message of the paper still holds.
The intertemporal preference of the agents at $t = 0$ is

$$V_0(c, y; \theta, \beta) = \left[ u_0(c_0) - h_0 \left( \frac{y_0}{\theta} \right) \right] + \beta \left[ \delta \left[ u_1(c_1) - h_1 \left( \frac{y_1}{\theta} \right) \right] + \delta^2 u_2(c_2) \right],$$

while at $t = 1$ it is

$$V_1(c, y; \theta, \beta) = u_1(c_1) - h_1 \left( \frac{y_1}{\theta} \right) + \beta \delta u_2(c_2),$$

à la quasi-hyperbolic discounting (Laibson, 1997). I will focus on the case where $\beta \in (0, 1)$, which measures the degree of present bias the agents suffer from. I will denote the time-consistent utility ($\beta = 1$) as $U_t(c, y; \theta)$.

Following O’Donoghue and Rabin (2001), non-sophisticated agents at $t = 0$ perceive their present bias in $t = 1$ to be $\hat{\beta} \in (\beta, 1]$. Let $W_1(c, y; \theta, \hat{\beta})$ denote the non-sophisticated agents’ perceived ex-post utility in $t = 1$:

$$W_1(c, y; \theta, \hat{\beta}) = u_1(c_1) - h_1 \left( \frac{y_1}{\theta} \right) + \hat{\beta} \delta u_2(c_2).$$

If $\hat{\beta} = 1$, the agent is fully naïve and unaware of the future-self’s present bias. If $\hat{\beta} = \beta$, the agent is sophisticated and fully aware of the bias. Partially naïve agents know they have present bias, $\bar{\beta} < 1$, but $\hat{\beta} > \beta$. In essence, they underestimate the severity of the future-self’s bias. Non-sophistication refers to $\hat{\beta} \in (\beta, 1]$ [15] I will refer to $\hat{\beta}$ as sophistication.

Agents vary in present bias, $\beta$, and sophistication $\hat{\beta}$, so types are represented by $(\theta_m, \beta, \hat{\beta})$, where $\hat{\beta} \in [\beta, 1]$ and $\beta \in [\beta, \bar{\beta}]$ with $\bar{\beta} \geq \beta$. In Section 4, I will analyze the case when $\bar{\beta} < 1$ and Section 5 will analyze the case when $\bar{\beta} = 1$. Let $\delta = 1$, which does not affect the results of the paper.

The timing is as follows: Before $t = 0$, the government designs the tax system, and has full commitment. The agents learn about their productivity at $t = 0$, and proceed to work, consume and save for each $t < 2$. The agents retire in $t = 2$.

The government tries to help the agents commit to the time-consistent counterpart, $U_0(c, y; \theta) = V_0(c, y; \theta, \beta = 1)$. I assume the non-sophisticated agents do not draw any inferences from the policies the government enacts, because they do not share the same prior as the government and are dogmatic in their beliefs. The government maximizes the following welfare criterion at $t = 0$

$$\sum_{m=1}^{M} \pi_m U_0(c_m, y_m; \theta_m),$$

____

14 Off-equilibrium threats and promises do not have bite if $\beta = 0$. The main idea of utilizing time inconsistency to raise welfare does not change if $\beta > 1$.

15 If $\hat{\beta} < \beta$, it is still possible for the government to take advantage of the incorrect belief.
where \((c_m, y_m)\) denotes the vector of allocations a \(\theta_m\) agent consumes. Since \(U\) is strictly concave in consumption, the government has a desire to insure agents against the realization of \(\theta\). I assume the government has no external revenue needs, so the feasibility constraint is

\[
\sum_{t=0}^{2} \sum_{m=1}^{M} \pi_m (y_{m,t} - c_{m,t}) = 0, \quad (2)
\]

with \(y_{m,2} = 0\) for all \(\theta_m \in \Theta\).

Finally, there are no private markets to insure against productivity shocks and no markets for illiquid assets or other commitment devices.

### 2.1 The Benchmarks

#### 2.1.1 No Private Information

In the no private information case, the government maximizes social welfare \((1)\) subject to the feasibility constraint \((2)\).

**Proposition 1** The efficient allocation \(\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}\) satisfies \((3)\) and for any \(\theta_m \in \Theta\):

(i.) full insurance: for any \(t\), \(c^*_m, t = c^*_t\), (ii.) consumption smoothing: for any \(t > 0\), \(u'_0(c^*_m, 0) = u'_t(c^*_m, t)\), and (iii.) efficient output: for any \(t < 2\), \(u'_t(c^*_m, t) = \frac{1}{\theta_m} h'_t(y^*_m, \theta_m)\).

With complete information, the government achieves full insurance regardless of present bias or sophistication. This is because the agents work according to their productivity. The government then chooses an appropriate linear savings subsidy to correct the distortion caused by the present bias. To see this, consider an income tax \(T^\beta_{m,t}\) for agents of productivity \(\theta_m\) and present bias \(\beta\) at \(t < 2\). Let \(y_t = y^*_{m,t} - T^\beta_{m,t}\) denote the after-tax income, which is the same for all productivity due to full insurance. Consider a savings subsidy of \(\tau^* = (\tau^*_1, \tau^*_2)\).

At \(t = 1\), for any savings \(s_1\) made at \(t = 0\), the agents solve

\[
\max_{c_1, s_2, c_2} u_1(c_1) - h_1 \left( \frac{y^*_{m,1}}{\theta_m} \right) + \beta u_2(c_2)
\]

subject to \(c_1 + s_2 \leq y_1 + (1 + \tau^*_1) s_1\) and \(c_2 \leq (1 + \tau^*_2) s_2\). If \(\tau^*_2 = \frac{1}{\beta} - 1\), then \(u'_1(c_1) = u'_2(c_2)\).

Let \(c_t(s_1)\) denote the optimal consumption given \(s_1\). By backward induction, at \(t = 0\), the agents solve

\[
\max_{c_0, s_1} u_0(c_0) - h_0 \left( \frac{y^*_{m,0}}{\theta_m} \right) + \beta \left[ u_1(c_1(s_1)) - h_1 \left( \frac{y^*_{m,1}}{\theta_m} \right) + u_2(c_2(s_1)) \right]
\]

\(^{16}\) Much of the literature on dynamically inconsistent preferences have evaluated welfare with the time-consistent utility.
subject to \( c_0 + s_1 = y_0 \). Similarly, the government can choose \( \tau^*_t \) such that \( u'_0(c_0) = u'_t(c_t) \) for any \( t > 0 \). It can then adjust the income tax so the efficient consumption is implemented.

### 2.1.2 Private Information without Time Inconsistency

With private information, the implementable allocations must be incentive compatible. The government maximizes social welfare \( (1) \) subject to the feasibility constraint \( (2) \) and the incentive compatibility constraints, \( \forall \theta_m, \theta_{m'} \in \Theta \),

\[
U_0(c_m, y_m; \theta_m) \geq U_0(c_{m'}, y_{m'}; \theta_m),
\]

which are evaluated at \( U_0 \) for time-consistent agents. Due to \( (3) \), the government implements the constrained efficient optimum.

**Proposition 2** The constrained efficient allocation \( \{(c_m^{**}, y_m^{**})\}_{\theta_m \in \Theta} \) satisfies \( (2), (3) \) and:

(i.) partial insurance: for any \( t \) and \( \theta_m, \theta_{m'} \in \Theta \), with \( \theta_m > \theta_{m'} \), \( c_{m,t}^{**} > c_{m',t}^{**} \); (ii.) consumption smoothing: for any \( t \) and \( \theta_m \in \Theta \), \( u'_t(c_{m,t}^{**}) = u'_{t+1}(c_{m,t+1}^{**}) \); and (iii.) output distortions: for any \( t < 2 \) and \( \theta_m < \theta_M \), \( u'_t(c_{m,t}^{**}) > \frac{1}{\theta_m} h'_t(y_{m,t}^{**} / \theta_m) \).

The constrained efficient allocation distorts the labor decisions of all agents except for the most productive agents \( \theta_M \). This distortion relaxes the incentive compatibility constraint, which allows the government to provide partial insurance. Hence, Proposition 2 characterizes the optimal trade-off between efficiency and equity.\(^{17}\)

### 3 The General Mechanism

Enlarged menus have been used to exploit time-inconsistent agents in the literature, see Esteban and Miyagawa (2005), Eliaz and Spiegler (2006) and Galperti (2015). In this section, I will introduce the betting mechanism and the conditional commitment mechanism for known bias and sophistication.

For a given present bias, \( \beta \), and sophistication, \( \hat{\beta} \), an enlarged menu \( C_m \) for type \( \theta_m \) agents is defined as \( C_m = \{(c_m, y_m), (c'_m, y'_m), \ldots\} \). An agent is assigned a menu \( C_m \) after reporting \( \theta_m \). Let \( (c^R_m, y^R_m) \in C_m \) be the real allocation, which is the optimal allocation the government implements. The government posts \( C = \{C_1, \ldots, C_M\} \). The agents then choose a menu \( C_m \) from \( C \) after learning \( \theta \) at \( t = 0 \). Without loss of generality, for all \( \theta_m \in \Theta \), set all

\(^{17}\)For more on the characterization of the constrained efficient allocation, see Hellwig (2007).
allocations in $t = 0$ equal to $\left( c_{m,0}^R, y_{m,0}^R \right)$, so agents start facing extraneous options at $t = 1$:

$$C_m = \left\{ \left( (c_{m,0}^R, y_{m,0}^R), (c_{m,1}^R, y_{m,1}^R), c_{m,2}^R \right), \left( (c_{m,0}^R, y_{m,0}^R), (c_{m,1}^R, y_{m,1}^R), c_{m,2}^R \right) \right\},$$

where the real allocations and extraneous options could be different, $(c_{m,1}^R, y_{m,1}^R) \neq (c_{m,1}^{'}, y_{m,1}^{'})$ and $c_{m,2}^R \neq c_{m,2}^{'},$ \(^{18}\)

Incentive compatibility is characterized by what the agent perceives the future-self will choose under both honest and dishonest reporting. Let

$$C_{\hat{\beta}}^m = \left\{ (c_m, y_m) \in C_m \mid (c_m, y_m) \in \arg \max_{(c_m, y_m) \in C_m} W_1(c_m', y_m'; \theta_m, \hat{\beta}) \right\}.$$

$C_{\hat{\beta}}^m$ denotes the set of allocations a truthful $\theta_m$ agent with sophistication $\hat{\beta}$ predicts the future-self would choose at $t = 1$. Let

$$C_{\hat{\beta}}^{m'|m} = \left\{ (c_{m'}, y_{m'}) \in C_{m'} \mid (c_{m'}, y_{m'}) \in \arg \max_{(c_{m'}, y_{m'}) \in C_{m'}} W_1(c_{m'}, y_{m'}'; \theta_m, \hat{\beta}) \right\}.$$

$C_{\hat{\beta}}^{m'|m}$ denotes the set of allocations a $\theta_m$ agent with sophistication $\hat{\beta}$ predicts the future-self would choose after misreporting to be a $\theta_{m'}$ agent. Incentive compatibility is thus expressed as, $\forall \theta_m, \theta_{m'} \in \Theta$,

$$\max_{(c_m, y_m) \in C_{\hat{\beta}}^m} V_0(c_m, y_m; \theta_m, \beta) \geq \max_{(c_{m'}, y_{m'}) \in C_{\hat{\beta}}^{m'|m}} V_0(c_{m'}, y_{m'}; \theta_m, \beta). \quad (4)$$

The incentive compatibility constraints (4) ensure truthful reporting of productivity. Additional constraints are needed to ensure the real allocations are implemented at $t = 1$. The executability constraints are, $\forall \theta_m \in \Theta$,

$$\left( c_{m}^R, y_{m}^R \right) \in \arg \max_{(c_{m}, y_{m}) \in C_m} V_1(c_m, y_m; \theta_m, \beta). \quad (5)$$

If the executability constraints (5) hold, the agent would choose the real allocations in $t = 1$. With non-common priors, the mechanism has to consider the agents’ beliefs in the incentive compatibility constraints and the government’s beliefs in the executability constraints. Other allocations besides the real allocations are off-equilibrium path.

\(^{18}\)Non-sophisticated agents could potentially learn their present bias, as in Ali (2011). By delaying the appearance of the enlarged menu till $t = 1$, learning could be ignored.
3.1 The Betting Mechanism

Non-sophisticated agents mispredict their future behavior, so their reporting strategies reflect misguided expectations. The government implements a betting mechanism to exploit this incorrect belief.

**Definition 1** A direct betting mechanism for sophistication \( \hat{\beta} \in (\beta, 1] \) has a menu \( C = \{C_m\}_{\theta_m \in \Theta} \) with \( C_m = \{(c^R_m, y^R_m), (c^I_m, y^I_m)\} \) and \((c^R_m, y^R_m) \neq (c^I_m, y^I_m)\) for some \( \theta_m \) satisfying the fooling constraints: \( \forall \theta_m, \theta_m' \in \Theta, (c^I_m, y^I_m) \in C^\beta_m \) and \((c^I_m', y^I_m') \in C^\beta_m'|m\).

By Definition 1, the enlarged menu in a betting mechanism has the off-equilibrium allocation \((c^I_m, y^I_m) \in C_m\) for all productivity. Non-sophisticated agents of productivity \( \theta_m \) predict they would choose \((c^I_m, y^I_m)\) in \( t = 1 \), which, following Eliaz and Spiegler (2006) and Heidhues and Koszegi (2010), will be referred to as the imaginary allocation. However, the government intends the agents to choose allocation \((c^R_m, y^R_m)\). The fooling constraints ensure the imaginary allocations are more appealing than the real allocations for \( W_1(c, y; \hat{\theta}, \hat{\beta}) \), so the benefits of truth-telling would be evaluated under the imaginary allocations.

**Definition 2** An allocation \( \{(c^R_m, y^R_m)\}_{\theta_m \in \Theta} \) is truthfully implementable for present bias \( \beta \) by a direct betting mechanism for sophistication \( \hat{\beta} \in (\beta, 1] \), if there exists \( \{(c^I_m, y^I_m)\}_{\theta_m \in \Theta} \) such that the following are satisfied: (i.) incentive compatibility, and (ii.) executability.

By Definition 2, to implement the real allocations, the executability constraints have to hold, which require the real allocations to be chosen at \( t = 1 \). Definition 2 also requires the imaginary allocations to satisfy incentive compatibility. This is because by the fooling constraints, the betting mechanism incentivizes truth-telling at \( t = 0 \) through off-equilibrium path imaginary allocations. This relaxes the incentive compatibility constraints because a portion of information rents is loaded on allocations that would not be chosen. In Section 4 I show that if Assumption 1 holds, all of the information rents can be loaded onto the imaginary allocations.

The imaginary allocations are not required to satisfy the feasibility constraint. The government is certain about the degree of the naïveté and present bias of the agents, so it places no weight on a future where it honors the delivery of imaginary allocations. Another concern is that the agents may realize that the aggregate imaginary allocation violates feasibility and doubt the validity of the government’s promise. However, each agent is infinitesimally small, and though an agent believes the future-self would consume the imaginary allocation, the agent does not consider the belief and behavior of others.
3.2 Conditional Commitment Mechanism

For sophisticated agents, the government can design an off-equilibrium path option that exacerbates the present bias, which will be chosen only if an agent misreports productivity. This type of mechanism will be called a *conditional commitment mechanism* (Esteban and Miyagawa, 2005).

**Definition 3** A direct conditional commitment mechanism for sophisticated agents \( (\hat{\beta} = \beta) \) has

\[
C = \{ C_m \}_{\theta_m \in \Theta} \quad \text{with} \quad C_m = \{ (c^R_m, y^R_m), (c^T_m, y^T_m) \} \quad \text{and} \quad (c^R_m, y^R_m) \neq (c^T_m, y^T_m) \quad \text{for some} \quad \theta_m.
\]

By Definition 3, the enlarged menu in a conditional commitment mechanism has the off-equilibrium allocation \((c^T_m, y^T_m) \in C_m\). I will refer to \((c^T_m, y^T_m)\) as the threat allocation. Sophisticated agents have the correct belief about \(\beta\), so the executability constraints imply that truth-telling is evaluated at the real allocation and the threat constraints imply that misreports are evaluated at the threat.

**Definition 4** An allocation \( \{ (c^R_m, y^R_m) \}_{\theta_m \in \Theta} \) is truthfully implementable for present bias \(\beta\) by a direct conditional commitment mechanism when \(\hat{\beta} = \beta\) if there exists \( \{ (c^T_m, y^T_m) \}_{\theta_m \in \Theta} \) such that the following are satisfied: (i.) incentive compatibility, and (ii.) executability.

By Definition 4, the threat, \((c^T_m, y^T_m)\), for type \(\theta_m\) is designed such that, after preference reversal, a type \(\theta_m\) agent who reports truthfully at \(t = 0\) would never choose it in \(t = 1\) (by the executability constraint). However, an agent who misreported as \(\theta_m\) would (by the threat constraint). Definition 4 requires the threat allocations to satisfy incentive compatibility to deter the agents from misreporting. This helps relax the incentive compatibility constraints. In Section 4, I will show that if Assumption 1 is satisfied, the cost of misreporting can be made arbitrarily large when evaluated with \(V_0(c, y; \theta, \beta)\).

3.2.1 Conditional Commitment Mechanism for Non-Sophisticates

Fully naïve agents have to be fooled, since they do not respond to threats. While sophisticated agents have to be threatened, because they can never be fooled. Since partially naïve agents also have demand for commitment, they are also susceptible to threats.

**Definition 5** A direct conditional commitment mechanism for sophistication \(\hat{\beta} \in (\beta, 1)\) has a menu \(C = \{ C_m \}_{\theta_m \in \Theta}\) with \(C_m = \{ (c^R_m, y^R_m), (c^T_m, y^T_m) \} \quad \text{and} \quad (c^R_m, y^R_m) \neq (c^T_m, y^T_m) \quad \text{for some} \quad \theta_m \)

\[
\forall \theta_m, \theta_m' \in \Theta, \quad (c^T_{m'}, y^T_{m'}) \in C^\beta_{m'|m}, \quad \text{and} \quad (c^R_m, y^R_m) \in C^\beta_m.
\]
Definition 5 is similar to Definition 3, but with the extra requirement that \((c_{m}, y_{m}) \in C_{m}^{\hat{\beta}}\) for all productivity. This is because for \(\hat{\beta} \in (\beta, 1)\), the threats are evaluated using \(\hat{\beta}\) instead of the actual present bias \(\beta\). Therefore, the threat constraints are defined to ensure the agents perceive their future-selves choosing the real allocation when report is truthful and the threat allocation if they misreported. The constraint \((c_{m}, y_{m}) \in C_{m}^{\hat{\beta}}\) is redundant when agents are sophisticated, because it coincides with the executability constraint when \(\hat{\beta} = \beta\). The definition of truthfully implementable real allocations when agents are non-sophisticated and a conditional commitment mechanism is implemented is analogous to Definition 4.

4 Hidden Present Bias and Sophistication

In this section, I will present the first main result of the paper. Consider an economy where all agents are time-inconsistent, but vary in present bias \(\beta\) and sophistication \(\hat{\beta}\), so types are represented by \((\theta_{m}, \beta, \hat{\beta}) \in \Theta \times [\underline{\beta}, \overline{\beta}] \times [\beta, 1]\), where \(\overline{\beta} < 1\). I will show that if Assumption 1 is satisfied, the efficient allocation is implementable by combining the betting and conditional commitment mechanisms. I will refer to it as the hybrid mechanism.

**Definition 6** A direct hybrid mechanism for sophistication \(\hat{\beta} \in (\beta, 1)\) has a menu \(C = \{C_{m}\}_{\theta_{m} \in \Theta}\) with \(C_{m} = \{(c_{R}^{m}, y_{R}^{m}), (c_{I}^{m}, y_{I}^{m}), (c_{T}^{m}, y_{T}^{m})\}\) and \((c_{R}^{m}, y_{R}^{m}) \neq (c_{I}^{m}, y_{I}^{m})\) for some \(\theta_{m}\) satisfying the fooling constraints: \(\forall \theta_{m}, \theta_{m}' \in \Theta, (c_{I}^{m}, y_{I}^{m}) \in C_{m}^{\hat{\beta}}\) and \((c_{I}^{m}, y_{I}^{m}') \in C_{m'}^{\hat{\beta}}\), and \((c_{R}^{m}, y_{R}^{m}) \neq (c_{T}^{m}, y_{T}^{m})\) for some \(\theta_{m}\) satisfying the threat constraints: \(\forall \theta_{m}, \theta_{m}' \in \Theta, (c_{T}^{m}, y_{T}^{m}) \in C_{m'}^{\hat{\beta}}\) and \((c_{R}^{m}, y_{R}^{m}) \in C_{m'}^{\hat{\beta}}\).

By Definition 6, the government implements a hybrid mechanism by choosing a fixed target sophistication \(\hat{\beta}\) and designs both a betting and conditional commitment mechanisms for the targeted sophistication. The following definition defines a truthfully implementable hybrid mechanism for present bias \(\beta\).

**Definition 7** An allocation \(\{(c_{R}^{m}, y_{R}^{m})\}_{\theta_{m} \in \Theta}\) is truthfully implementable for present bias \(\beta\) by a direct hybrid mechanism for \(\hat{\beta} \in (\beta, 1)\) if there exists \(\{(c_{I}^{m}, y_{I}^{m})\}_{\theta_{m} \in \Theta}\) and \(\{(c_{T}^{m}, y_{T}^{m})\}_{\theta_{m} \in \Theta}\) such that incentive compatibility (4) and executability constraints (5) are satisfied.

The following Lemma analyzes the case for observable present bias. It shows that if both the betting and conditional commitment mechanisms for \(\hat{\beta} \in (\beta, 1)\) can implement the efficient allocation, then the hybrid mechanism implements the efficient allocation with hidden productivity and sophistication.
Lemma 1 For any \( \beta \in [\beta, \bar{\beta}] \), betting mechanisms implementing the efficient allocation for \( \hat{\beta} \in (\beta, 1) \) also implement it for any \( \hat{\beta}' \geq \hat{\beta} \). Similarly, for any \( \beta \in [\beta, \bar{\beta}] \), conditional commitment mechanisms implementing the efficient allocation for \( \beta \in (\beta, 1) \) also implement it for any \( \hat{\beta}' \leq \hat{\beta} \).

Proof Suppose \( \{(c'_m, y'_m)\}_{\theta_m \in \Theta} \) implements the efficient allocation for sophistication \( \hat{\beta} > \beta \). This implies that \( c'_{m,2} > c^*_{m,2} \) for any \( \theta_m > \theta_1 \). Hence, the fooling constraints are satisfied for any \( \hat{\beta}' > \hat{\beta} \), so it is incentive compatible. Notice for any \( \hat{\beta}' > \hat{\beta} \), the executability constraints are satisfied. Therefore, \( \{(c'_m, y'_m)\}_{\theta_m \in \Theta} \) also implements the efficient allocation for \( \hat{\beta}' > \hat{\beta} \).

Suppose \( \{(c'_m, y'_m)\}_{\theta_m \in \Theta} \) implements the efficient allocation for sophistication \( \hat{\beta} > \beta \). This implies \( c^*_2 > c^*_{m,2} \), because by the incentive compatibility and threat constraints:

\[
\begin{align*}
  u_2(c^*_2) - u_2(c^*_{m,2}) &> \left[ u_1(c^*_{m,1}) - h_1 \left( \frac{y^*_{m,1}}{\theta_{m+1}} \right) \right] - \left[ u_1(c'_1) - h_1 \left( \frac{y^{m+1}_{m+1}}{\theta_{m+1}} \right) \right] \\
  &\geq \left[ u_1(c^*_{m,1}) - h_1 \left( \frac{y^*_{m,1}}{\theta_{m+1}} \right) \right] - \left[ u_1(c'_1) - h_1 \left( \frac{y^{m+1}_{m+1}}{\theta_{m+1}} \right) \right] \\
  &\geq \hat{\beta} \left[ u_2(c^*_2) - u_2(c^*_{m,2}) \right].
\end{align*}
\]

Hence, the threat constraint is relaxed for any \( \hat{\beta}' < \hat{\beta} \). Since the executability constraint also holds, this implies the threat and executability constraints are satisfied for any \( \hat{\beta}' < \hat{\beta} \). Therefore, for \( \hat{\beta}' < \hat{\beta} \), incentive compatibility is satisfied and \( \{(c'_m, y'_m)\}_{\theta_m \in \Theta} \) implements the efficient allocation too. \( \blacksquare \)

By Lemma 1, for a known \( \beta \), a betting mechanism designed for sophistication \( \hat{\beta} \) agents can also fool agents who are more naïve. This is because fooling the less naïve agents is more difficult, so incentives that could screen the productivity of less naïve agents will also work for more naïve agents. Similarly, a conditional commitment mechanism designed for sophistication \( \hat{\beta} \) agents can also threaten agents who are more sophisticated. This is because threatening the less sophisticated agents is more difficult, so incentives that could separate the productivity of less sophisticated agents will also work for more sophisticated agents. The first main result of the paper shows that if Assumption 1 holds, then the efficient allocation is implementable for Mirrlees taxation with hidden present bias and sophistication by using a hybrid mechanism.

Theorem 1 For the environment where all agents are time-inconsistent with hidden present bias and sophistication, if Assumption 1 holds, the efficient allocation \( \{(c'_m, y'_m)\}_{\theta_m \in \Theta} \) is truthfully implementable with a hybrid mechanism.
Proof Set $(c^R_m, y^R_m) = (c^m_m, y^m_m); y^T_m = y^m_m, c^T_m,0 = c^*_m,0$ and $y^T_m,0 = y^*_m,0$ for all $\theta_m \in \Theta$. First, I will show that the betting mechanism can implement the efficient allocation for known present bias and sophistication $\hat{\beta} \in (\beta, 1)$ when Assumption 1 is satisfied. Notice the following programming problem has no finite solution: for all $\theta_m \in \Theta,$

$$\max_{c^m_{1,1}, c^m_{1,2}} u_1 (c^m_{1,1}) + \hat{\beta} u_2 (c^m_{1,2}),$$

subject to $u_1 (c^m_{1,1}) + \beta u_2 (c^m_{1,2}) \leq u_1 (c^*_m,1) + \beta u_2 (c^*_m,2).$ This follows from $\hat{\beta} > \beta$ and Assumption 1 which implies $u_2$ is unbounded above and $u_1$ is unbounded below. Therefore, the executability and fooling constraints can be satisfied. Finally, notice that local downward incentive compatibility implies global incentive compatibility. Since the programming problem above has no finite solution, the imaginary allocations can always be chosen so the incentive compatibility constraints are satisfied for any information rent.

Next, I will show that the conditional commitment mechanism can implement the efficient allocation for known present bias and sophistication $\hat{\beta} \in (\beta, 1)$ when Assumption 1 is satisfied. Notice the following programming problem has no finite solutions: for all $\theta_m \in \Theta,$

$$\min_{c^m_{1,1}, c^m_{1,2}, y^T_m} u_1 (c^m_{1,1}) - h_1 \left( \frac{y^T_m}{\theta_m} \right) + \beta u_2 (c^m_{1,2}),$$

subject to

$$u_1 (c^m_{1,1}) - h_1 \left( \frac{y^T_m}{\theta_{m+1}} \right) + \hat{\beta} u_2 (c^m_{1,2}) \geq u_1 (c^*_m,1) - h_1 \left( \frac{y^*_m,1}{\theta_{m+1}} \right) + \beta u_2 (c^*_m,2). \quad (6)$$

This is because $h_1$ is strictly increasing, strictly convex and $\theta_{m+1} > \theta_m,$ so $y^T_m$ can be increased arbitrarily. Also, by Assumption 1 $u_1$ is unbounded above, so $c^T_m$ can be increased to compensate for the increase in $y^T_m$ so that (6) is satisfied. Note that if $y^T_m,1 > y^*_m,1$ and (6) holds, then the threat allocation is more appealing for all agents more productive than $\theta_{m+1}.$ Hence, the threat and executability constraints hold.

Suppose (6) holds and by Proposition 1 $c^*_{m,t} = c^*_{m+1,t} = c^*_t,$ the local downward incentive compatibility constraint for $\theta_{m+1}$ can be written as

$$\left( 1 - \hat{\beta} \right) [u_2 (c^*_2) - u_2 (c^T_m)] \geq \frac{1}{\beta} \left[ h_0 \left( \frac{y^*_m,0}{\theta_{m+1}} \right) - h_0 \left( \frac{y^*_m,0}{\theta_{m+1}} \right) \right] + h_1 \left( \frac{y^*_m,1}{\theta_{m+1}} \right) - h_1 \left( \frac{y^*_m,1}{\theta_{m+1}} \right),$$

because $u_2$ is unbounded below and strictly increasing, there exists $L_{m,2}$ such that all $c_2 \leq L_{m,2}$ satisfies the local incentive compatibility constraint for $\theta_{m+1}.$ Trivially, $c^T_m$ can be adjusted
so the threat and executability constraints still hold, because \( u_1 \) is unbounded above. Finally, local downward incentive compatibility implies global incentive compatibility.

Finally, I will show that with hidden present bias and sophistication, the hybrid mechanism can implement the efficient allocation. Construct \( \{(c^I_m, y^I_m)\}_{\theta_m \in \Theta} \) to implement the efficient allocation for some \( \hat{\beta} \in (\overline{\beta}, 1) \) and \( \beta = \overline{\beta} \). By Lemma 1, agents with \( \hat{\beta}' > \hat{\beta} \) and \( \beta = \overline{\beta} \) would be fooled. Only the executability constraint depends on \( \beta \). Since \( c^I_{m,2} > c^*_2 \), then for any \( \beta < \overline{\beta} \), the executability constraints also hold. Therefore, \( \{(c^I_m, y^I_m)\}_{\theta_m \in \Theta} \) implements the efficient allocation for more naïve and present-biased agents.

Construct \( \{(c^T_m, y^T_m)\}_{\theta_m \in \Theta} \) to implement the efficient allocation for \( \hat{\beta} \) and \( \beta = \overline{\beta} \). For all \( \theta_m \), fix \( (c^I_{m,1}, c^I_{m,2}) \) and choose \( (c^T_{m,1}, c^T_{m,2}) \) such that

\[
 u_1(c^I_{m,1}) - h_1\left(\frac{y^I_{m,1}}{\theta_{m+1}}\right) + \hat{\beta} u_2(c^I_{m,2}) = u_1(c^I_{m,1}) - h_1\left(\frac{y^I_{m,1}}{\theta_{m+1}}\right) + \hat{\beta} u_2(c^T_{m,2}).
\]

I have shown that it is possible to construct \( \{(c^T_m, y^T_m)\}_{\theta_m \in \Theta} \) to implement the efficient allocation for \( \hat{\beta} \) and \( \beta = \overline{\beta} \). Inequality (7) ensures agents who misreport would be indifferent between the threat and imaginary allocations. Note that this does not change the reporting strategy for \( \hat{\beta}' > \hat{\beta} \) in a hybrid mechanism. By Lemma 1, \( \{(c^T_m, y^T_m)\}_{\theta_m \in \Theta} \) implements the efficient allocation for agents with \( \hat{\beta}' < \hat{\beta} \) and \( \beta = \overline{\beta} \). Finally, only the executability constraints depend on \( \beta \). For any \( \beta > \overline{\beta} \), the executability constraint for threat allocations are relaxed. Hence, the hybrid mechanism implements the efficient allocation.

Theorem 1 depends on three key observations. First, the efficient allocation is implementable with betting and conditional commitment mechanisms for a known present bias and sophistication when Assumption 1 holds. Therefore, the betting and conditional mechanisms facilitate the government in screen productivity. Secondly, the government can design betting and conditional commitment mechanisms for a target sophistication to screen sophistication. This is because by Lemma 1, agents more naïve than the target would self-select into the betting mechanism, while agents less naïve than the target would self-select into the conditional commitment mechanism. Finally, to screen present bias, the betting mechanism needs to be designed for the least present-biased agent while the conditional commitment mechanism needs to be tailored to the most present-biased agent. This is because if the real allocation is executable for the least present-biased agent in a betting mechanism, then it is also executable for more present-biased agents. Similarly, if the real allocation is executable for the most present-biased agent in a conditional commitment mechanism, then it is also executable for less present-biased agents. As a result, the hybrid mechanism screens
productivity, present bias and sophistication costlessly when Assumption 1 holds. It is also important to point out that Theorem 1 is robust to changes in the joint distribution of \((θ_m, β, \hat{β})\). In addition to the primitives introduced in Section 2 the government does not need to know more than \(β\) and \(\hat{β}\).

To see how the betting and conditional commitment mechanisms screen productivity and why Assumption 1 is a sufficient condition for implementing the efficient allocation, consider an economy with two productivity types \(θ = \{θ_L, θ_H\}\), where \(θ_H > θ_L\), and fixed present bias \(β\). Let \(\{(c^R_m, y^R_m)\}_{θ_m ∈ Θ}\) be the efficient allocation, where \(c^R_{H,t} = c^R_{L,t} = c^*_t\) and \(y^R_{m,t} = y^*_m,t\) for all \(θ_m ∈ Θ\). The efficient allocation satisfies Proposition 1 so \(y^*_H > y^*_L\). For simplicity, I will demonstrate the betting mechanism for the fully naïve case: \(\hat{β} = 1\), and the conditional commitment mechanism for the fully sophisticated case \(\hat{β} = β\).

I will first examine the betting mechanism. Set \(c^I_{m,0} = c^*_0\) and \(y^I_{m,t} = y^*_m,t\) for all \(θ_m\). In Figure 1, the flatter solid (blue) curve represents the indifference curve from the perspective of \(t = 0\) at allocation \((c^*_1, c^*_2)\). The present-biased agents value \(c_2\) less at \(t = 1\) than at \(t = 0\), so the steeper solid (red) curve represents the indifference curve from the perspective of \(t = 1\) at allocation \((c^*_1, c^*_2)\). The imaginary allocations have to be in the area bounded by the solid indifference curves in the north-west region: below the red curve and above the blue curve. This is because for the efficient allocations to be implemented, the agents have to prefer it over the imaginary allocations. This would ensure the executability constraints are satisfied.

Furthermore, the incentive compatibility constraints provide upper and lower bounds to the difference in utility between the two types of agents evaluated at the imaginary

\[\begin{align*}
\text{Figure 1: Finding the Imaginary Allocations}
\end{align*}\]
allocations. In essence,

\[
\sum_{t=0}^{1} \beta^t \left[ h_t \left( \frac{y_{H,t}^*}{\theta_L} \right) - h_t \left( \frac{y_{L,t}^*}{\theta_L} \right) \right] \geq \beta \left[ u_1(c^I_{H,1}) + u_2(c^I_{H,2}) \right] - \beta \left[ u_1(c^I_{L,1}) + u_2(c^I_{L,2}) \right] \\
\geq \sum_{t=0}^{1} \beta^t \left[ h_t \left( \frac{y_{H,t}^*}{\theta_H} \right) - h_t \left( \frac{y_{L,t}^*}{\theta_H} \right) \right],
\]

where the upper bound is derived from the incentive compatibility constraint for \(\theta_L\) and the lower bound is derived from the incentive compatibility constraint for \(\theta_H\). Figure 1 shows that the difference in utility between the \(\theta_H\) and \(\theta_L\) imaginary allocations have to be greater than the thin dashed lines (the lower bound) and less than the thick dashed lines (the upper bound) for incentive compatibility to be satisfied.

When Assumption 1 is satisfied, the indifference curves never touch the axis. Hence, it would always be possible to find imaginary allocations that satisfy the incentive compatibility, fooling and executability constraints by increasing \(c^I_{H,2}\) and decreasing \(c^I_{L,2}\). In other words, if Assumption 1 is satisfied, the government can always decrease consumption in \(t = 1\) and load the information rent on retirement consumption to simultaneously satisfy both incentive compatibility and executability.

Next, to see how the conditional commitment mechanism implements the efficient allocation for sophisticated agents when Assumption 1 holds, set \(c^T_{L,0} = c^*_0\) and \(y^T_{L,0} = y^*_L\), so the threat occurs at \(t = 1\). Let \(\Phi^j_{i,k} = u_1(c^j_{i,1}) - h_1 \left( \frac{y^j_{i,1}}{\theta_k} \right) \) and \(\Phi^j_{k,k} = \Phi^1_k\), where \(i \in \{R, T\}\) and \(j, k \in \{L, H\}\). This demonstration will proceed in two steps. For the first step, I will first show how to construct threat allocations to deter misreporting. I will then show how it can be adjusted so that truthful agents would never choose it. From incentive compatibility, threat and executability constraints and by \(\theta_H > \theta_L\) and \(\beta < 1\), the efficient and threat allocations have to satisfy: \(\Phi^T_{L,H} > \Phi^R_{L,H} > \Phi^R_{L} > \Phi^T_{H}\), and \(c^*_2 > c^T_{L,2}\).

Figure 2 shows how the incentive compatibility constraint restricts the set of threat allocations. The steeper solid (red) curve represents the indifference curve from the perspective of \(t = 1\) for the \(\theta_H\) agent who pretended to be \(\theta_L\) in \(t = 0\). The flatter solid (blue) curve represents the indifference curve from the perspective of \(t = 1\) for the \(\theta_H\) agent who reported truthfully in \(t = 0\). The dashed (red) curve represents the indifference curve from the perspective of \(t = 0\) for the truthful \(\theta_L\) agent. Figure 2 shows when Assumption 1 holds, the government can choose \((c^T_L, y^T_L)\) such that the incentive compatibility constraint is satisfied by decreasing \(c^T_{L,2}\) and increasing \(\Phi^T_{L,H}\). Furthermore, when Assumption 1 holds, it is possible to increase \(c^T_{L,1}\) so that the threat constraint holds for any arbitrarily small \(c^T_{L,2}\).
Finally, I will show that \((c_{T,1}^*, y_{T,1}^*)\) can be chosen so that the executability constraint is satisfied. To see this, fix the choice of \(c_{T,2}^*\) and \(\Phi_{T,L,H}^*\) at the level shown in Figure 2. If the threat satisfies incentive compatibility, threat and executability constraints, it implies

\[
\Delta u_2 \geq \Phi_{L,H}^T - \Phi_{L,H}^R + \frac{1}{\beta} \Delta h_0 > \Phi_{L,H}^T - \Phi_{L,H}^R \geq \beta \Delta u_2 \geq \Phi_{L,H}^T - \Phi_{L}^R,
\]

where \(\Delta u_2 \equiv [u_2(c_2^*) - u_2(c_{T,2}^*)]\) and \(\Delta h_0 = h_0(y_{L,1}^*/\theta_H) - h_0(y_{L,1}^*/\theta_H)\). The problem now is to find \(c_{T,1}^*\) and \(y_{T,1}^*\) such that \(u_1(c_{T,1}^*) - h_1(y_{L,1}^*/\theta_H) = \Phi_{L,H}^T\) and satisfies the executability constraint, \(\beta \Delta u_2 \geq \Phi_{L}^T - \Phi_{L}^R\).

In Figure 3, the flatter thick solid (blue) curve represents the indifference curve of \(\Phi\) for the \(\theta_H\) agents at allocation \((c_{T,1}^*, y_{L,1}^*)\). The steeper solid (red) curve represents the indifference curve of \(\Phi\) for the \(\theta_L\) agents at allocation \((c_{T,1}^*, y_{L,1}^*)\). The dashed (blue) curve represents the indifference curve of \(\Phi\) for the \(\theta_H\) agent at allocation \((c_{T,1}^*, y_{T,1}^*)\), chosen so that \(u_1(c_{T,1}^*) - h_1(y_{L,1}^*/\theta_H) = \Phi_{L,H}^T\) and the executability constraint holds. In essence, the government can increase \(y_{L,1}\) to discourage \(\theta_L\) agents from choosing the threat allocation. While by Assumption 1, it can increase \(c_{T,1}\) the threat remains potent for \(\theta_H\) agents. Hence, Assumption 1 provides the government sufficient leverage to discipline the agents through off-equilibrium path policies.
5 Model with Time-Consistent Agents

In this section, I will discuss the consequences of introducing time-consistent (TC) agents. The government does not observe whether agents are time-inconsistent (TI) or TC. With TC agents present, the efficacy of betting or conditional commitment mechanisms is limited. TC agents cause distortions, because they follow through with their consumption plans, so off-path policies are ineffective and the efficient optimum is no longer attainable. In this section, I also present the second main result of the paper, which shows that welfare increases monotonically with the proportion of time-inconsistent agents in the economy. The proofs for this section are in the online appendix.

I will focus on TI agents with the same present bias $\beta$ and sophistication $\hat{\beta} < 1$. With Assumption\footnote{A discussion of fully naïve agents is provided in Section 7.} this environment is the same as a setting with heterogeneous present bias and sophistication where the most present-biased agent has bias $\beta$ and the least sophisticated agent has sophistication $\hat{\beta}$. I have excluded fully naïve agents in this analysis, because separation of TC and fully naïve agents is not possible at $t = 0$\footnote{A discussion of fully naïve agents is provided in Section 7.} The government is uncertain whether the agents are TC ($\beta = 1$) or TI ($\beta < 1$), with probability $\Pr(TI) = \phi$. The TC agents know their consistency, while TI agents could be non-sophisticated. I assume the distribution of productivity is independent of the agents’ consistency. I will first present the result for conditional commitment mechanisms and then discuss betting mechanisms.
5.1 Conditional Commitment Mechanisms

If all TI agents are not fully naïve ($\hat{\beta} < 1$), then it is possible to separate TI agents from TC agents. If the government implements a conditional commitment mechanism, then it can design the following menu for TC agents:

$$C_{m}^{TC} = \left\{ (c_{m}^{P}, y_{m}^{P}); (c_{m}^{D}, y_{m}^{D}) \right\}$$

and the following menu for TI agents: $C_{m}^{TI} = \left\{ (c_{m}^{R}, y_{m}^{R}); (c_{m}^{T}, y_{m}^{T}) \right\}$. The allocation $(c_{m}^{P}, y_{m}^{P})$ is the persistent allocation, and it is the allocation the government implements for $\theta_{m}$ TC agents. The allocation $(c_{m}^{D}, y_{m}^{D})$ is referred to as the deterrent allocation, and it is meant to deter the TI agents from misreporting as $\theta_{m}$ TC agents. The idea is similar to Galperti (2015), where unused options were introduced to deter time-inconsistent agents from mimicking time-consistent agents. I will show that the optimal allocations in this environment will always make the presence of deterrent allocations necessary. The deterrent allocations work in a similar fashion as threat allocations. The government offers $C = \{C_{m}^{TC}, C_{m}^{TI}\}_{\theta_{m} \in \Theta}$. The mechanism is meant to separate agents along two dimensions: productivity and consistency.

Let

$$C_{m'}^{\hat{\beta}|m} = \left\{ (c_{m'}, y_{m'}) \in C_{m'}^{TC} \middle| (c_{m'}, y_{m'}) \in \arg \max_{(c_{m'}, y_{m'}) \in C_{m'}^{TC}} W_{1}(c_{m'}, y_{m'}; \hat{\theta}_{m}, \hat{\beta}) \right\}.$$  

$C_{m'|m}^{\hat{\beta}}$ denotes the set of allocations a TI agent with productivity $\theta_{m}$ and sophistication $\hat{\beta}$ predicts the future-self will choose in $t = 1$ after misreporting to be TC with productivity $\theta_{m'}$. Let

$$C_{m}^{1} = \left\{ (c_{m}, y_{m}) \in C_{m}^{TC} \middle| (c_{m}, y_{m}) \in \arg \max_{(c_{m}, y_{m}) \in C_{m}^{TC}} U_{0}(c_{m}', y_{m}'; \theta_{m}) \right\}.$$  

$C_{m}^{1}$ denotes the set of allocations a truth-telling TC agent would choose. Let $C_{m}^{\hat{\beta}}$ and $C_{m'|m}^{\hat{\beta}}$ be defined as before. The following definition defines a direct conditional commitment mechanism with TC agents.

**Definition 8** A direct conditional commitment mechanism for sophistication $\hat{\beta} \in (\beta, 1)$ with TC agents has $C = \{C_{m}^{TC}, C_{m}^{TI}\}_{\theta_{m} \in \Theta}$, with $C_{m}^{TC} = \{ (c_{m}^{P}, y_{m}^{P}); (c_{m}^{D}, y_{m}^{D}) \}$ and $C_{m}^{TI} = \{ (c_{m}^{R}, y_{m}^{R}); (c_{m}^{T}, y_{m}^{T}) \}$ satisfying: (i.) threat constraints: $\forall \theta_{m}, \theta_{m'} \in \Theta$, $(c_{m}^{R}, y_{m}^{R}) \in C_{m}^{\hat{\beta}}$ and $(c_{m'}^{T}, y_{m'}^{T}) \in C_{m'|m}^{\hat{\beta}}$, (ii.) deterrent constraints: $\forall \theta_{m}, \theta_{m'} \in \Theta$, $(c_{m}^{P}, y_{m}^{P}) \in C_{m}^{1}$, and $(c_{m'}^{D}, y_{m'}^{D}) \in C_{m'|m}^{\hat{\beta}}.$
The incentive compatibility constraints for the TI agents are, \( \forall \theta_m, \theta_{m'} \in \Theta, \)
\[
\max_{(c_m, y_m) \in C^\text{\text{TC}}_{m'}} V_0(c_m, y_m; \theta_m, \beta) \leq \max \left\{ \max_{(c_{m'}, y_{m'}) \in C^{\text{\text{TC}}}_{m'|m}} V_0(c_{m'}, y_{m'}; \theta_m, \beta) \right\},
\]
By (8), it is optimal for the TI agents to report truthfully about their productivity and consistency. Let
\[
C^{1\mid \beta}_{m'|m} = \left\{ (c_{m'}, y_{m'}) \in C^{\text{\text{TC}}}_{m'} \mid (c_{m'}, y_{m'}) \in \arg \max_{(c_{m'}, y_{m'}) \in C^{\text{\text{TC}}}_{m'}} U_0(c_{m'}, y_{m'}; \theta_m) \right\},
\]
\[
C^{1}_{m'|m} = \left\{ (c_{m'}, y_{m'}) \in C^{\text{\text{TC}}}_{m'} \mid (c_{m'}, y_{m'}) \in \arg \max_{(c_{m'}, y_{m'}) \in C^{\text{\text{TC}}}_{m'}} U_0(c_{m'}, y_{m'}; \theta_m) \right\}.
\]
Hence, \( C^{1\mid \beta}_{m'|m} \) denotes the set of allocations a TC agent with productivity \( \theta_m \) would select from a menu for TI agents with productivity \( \theta_{m'} \), while \( C^{1}_{m'|m} \) denotes the set of allocations a TC agent would select if productivity was misreported as \( \theta_{m'} \) and was truthful about consistency. The incentive compatibility constraints for the TC agents are, \( \forall \theta_m, \theta_{m'} \in \Theta, \)
\[
\max_{(c_m, y_m) \in C^{1}_{m'}} U_0(c_m, y_m; \theta_m) \geq \max \left\{ \max_{(c_{m'}, y_{m'}) \in C^{1}_{m'|m}} U_0(c_{m'}, y_{m'}; \theta_m) \right\}, \tag{9}
\]
Incentive compatibility constraints (9) discourage the TC agents from misreporting productivity or consistency. The executability constraints are defined by (5).

**Definition 9** The allocation \( \{(c^P_m, y^P_m), (c^R_m, y^R_m)\}_{\theta_m \in \Theta} \) is truthfully implementable for present bias \( \beta \) by a direct conditional commitment mechanism for sophistication \( \hat{\beta} \) with TC agents if there exists \( \{(c^P_m, y^P_m), (c^T_m, y^T_m)\}_{\theta_m \in \Theta} \) such that (i.) incentive compatibility, and (ii.) executability are satisfied.

The government maximizes welfare
\[
\sum_{\theta_m \in \Theta} \pi_m [\phi U_0(c^R_m, y^R_m; \theta_m) + (1 - \phi) U_0(c^P_m, y^P_m; \theta_m)], \tag{10}
\]
subject to the incentive compatibility constraints, executability constraints, credible threat constraints, deterrent constraints and the feasibility constraint

\[
\sum_{\theta_m \in \Theta} \left\{ \phi \pi_m \left[ \sum_{t=0}^{2} (y^R_{m,t} - c^R_{m,t}) \right] + (1 - \phi) \pi_m \left[ \sum_{t=0}^{2} (y^P_{m,t} - c^P_{m,t}) \right] \right\} = 0, \quad (11)
\]

where \( y_2 = 0 \). The threat and deterrent allocations can help relax (8). As a result, the government only needs to deter misreporting from the TC agents. The following theorem shows how the government takes advantage of the TI agents.

**Theorem 2** For a conditional commitment mechanism, if Assumption 1 holds and \( \phi \in (0, 1) \) and \( \hat{\beta} \in (\beta, 1) \), then there exists \( \bar{\theta} > \theta_1 \) such that (i.) for \( \theta_m \geq \bar{\theta} \), \( c^P_m \geq c^R_m \) with \( y^R_m > y^P_m \), (ii.) for \( \theta_1 < \theta_m < \bar{\theta} \), \( (c^P_m, y^P_m) < (c^R_m, y^R_m) \), (iii.) \( (c^P_1, y^P_1) = (c^R_1, y^R_1) \).

By Theorem 2, full insurance is no longer incentive compatible when TC agents are in the economy. The high productivity \( (\theta_m \geq \bar{\theta}) \) TC agents require information rents, so they have lower marginal utilities from consumption and lower disutility from effective labor \( y \).

When Assumption 1 holds, the incentive compatibility constraints of TI agents are non-binding. The government exploits the higher productivity TI agents by requiring them to work more and consume less, which increases the resources available for redistribution. As a result, TC agents with \( \theta_m \geq \bar{\theta} \) would never misreport to be TI agents of the same productivity. For agents with lower productivity \( (\theta_m < \bar{\theta}) \), the government would exploit the TI agents by requiring them to work more, but also compensate them with more consumption for insurance. The consumption is limited by the incentive compatibility constraint for the higher productivity agents. This increase in production more than offsets the increase in consumption, so it also increases the resources available for redistribution. The government refrains from exploiting the TI agents with the lowest productivity by bunching them with the TC agents, because any exploitation would only lead to less insurance. The only binding incentive compatibility constraints are the downward adjacent incentive compatibility constraints for the TC agents. For \( \theta_m > \theta_1 \), TI agents have strictly lower lifetime utility than TC agents of the same productivity. As a result, deterrent allocations must always be present in the menu for TC agents.

To see how deterrent allocations are constructed, for any given \( \{(c^R_m, y^R_m)\}_{\theta_m \in \Theta} \) and \( \{(c^P_m, y^P_m)\}_{\theta_m \in \Theta} \), let \( (c^D_m, y^D_m) \) with \( y^D_m = y^P_m \), \( c^D_{m,0} = c^P_{m,0} \) and \( c^D_{m,t} = c^D_t \) for all \( t > 0 \) be the deterrent allocation satisfying:

\[
\min_{\theta_m, \theta'_m \in \Theta} W_1 \left( c^D_m, y^D_m; \theta_m, \hat{\beta} \right) \geq \max_{\theta_m, \theta'_m \in \Theta} W_1 \left( c^P_{m'}, y^P_{m'}; \theta_m, \hat{\beta} \right), \quad (12)
\]
and

\[
\min_{\theta_m \in \Theta} V_0 (c^R_m, y^R_m; \theta_m, \beta) \geq \max_{\theta_m, \theta'_m \in \Theta} V_0 (c^D_m', y^D_m'; \theta_m, \beta). \tag{13}
\]

By inequality \(12\), any TI agent who misreports as TC would select the deterrent allocation over the persistent allocation. Inequality \(13\) guarantees the TI agents would prefer to report their consistency truthfully. If \(13\) is satisfied, the TC agents would never choose the deterrent allocations over the persistent allocations. If \(12\) and \(13\) hold, then TI agents of any productivity would never misreport to be TC agents. Finally, the deterrent allocations can always be constructed such that \(12\) and \(13\) are satisfied.

Let the intertemporal wedge be defined as
\[
\tau_C^t = 1 - \frac{u_t'(c_t)}{u_{t+1}'(c_{t+1})},
\]
and the labor wedge as
\[
\tau_L^t = 1 - \frac{\frac{1}{\beta} k'_t(y_t)}{u_t'(c_t)}.
\]
The following theorem characterizes the optimal allocation and wedges in an environment with TC agents.

**Theorem 3** For a conditional commitment mechanism, if Assumption 1 holds and \(\hat{\beta} \in (0, 1)\) and \(\hat{\beta} \in (\beta, 1)\), then the optimal allocation has the following properties: for any \(t < 2\), (i.) \(\tau_C^t = 0\) for all agents, (ii.) \(\tau_L^t = 0\) for all TI agents with \(\theta_m \geq \hat{\theta}\) and \(\theta_M\) TC agents, (iii.) \(\tau_L^t \geq 0\) for all TI agents with \(\theta_m < \hat{\theta}\) and \(\tau_L^t > 0\) for all TC agents with \(\theta_m < \theta_M\), and (iv.) \(c^R_m = c^R_m'\) for all TI agents with \(\theta_m \geq \hat{\theta}\).

The usual trade-off between insurance and output efficiency is present in this economy. Theorem 3 demonstrates how the output of the less productive agents is distorted downwards. This is standard in Mirrlees taxation. Also, the government is able to provide consumption smoothing for all agents, so the intertemporal wedge is undistorted.

The next corollary shows that as long as \(\hat{\beta} \in (\beta, 1)\), then the conditional commitment mechanism can implement the same optimal allocation for different levels of sophistication. This follows from the fact that the optimal allocation in a conditional commitment mechanism does not depend on the sophistication of the TI agents.

**Corollary 1** In a conditional commitment mechanism, the optimal allocation is the same for any sophistication \(\hat{\beta} \in [\beta, 1)\).

The constrained efficient optimum is achieved when no TI agents are present \((\phi = 0)\). Section 4 has shown that the efficient optimum is attainable when the economy is populated by TI agents \((\phi = 1)\). Let \(V^T(\phi)\) denote the welfare under a conditional commitment

---

20To see how, first choose \(c^D_m\) so \(13\) holds. Next, increase \(c^D_1\) and decrease \(c^D_2\) such that \(V_0 (c^D_m', y^D_m'; \theta_m, \beta)\) remains unchanged, and since \(1 > \hat{\beta}\) and Assumption 1 holds, then it is possible to find \(c^D_1\) and \(c^D_2\) such that \(12\) holds.
mechanism with measure $\phi$ of TI agents. The second main result of the paper shows that social welfare increases as the proportion of TI agents increases, as shown in Figure 4.

**Theorem 4** $W^T(\phi)$ increases continuously with $\phi$ from the constrained efficient optimum to the full information efficient optimum.

![Figure 4: Welfare and Proportion of Time-Inconsistent Agents](image)

As the mass of TI agents increases, the first order effect is an increase in the available resources for redistribution, which comes from the allocation patterns in Theorem 2. Theorem 2 shows that the optimal policy induces TI agents to produce more output and consume less. This relaxes the resource constraint. The second order effect is that with less TC agents, the government can provide each TC agent more information rent using fewer resources. This relaxes the incentive compatibility constraints for TC agents [9]. Hence, the government is able to provide better insurance with more TI agents.

### 5.2 Betting Mechanisms

The government can also implement a betting mechanism when $\hat{\beta} \in (\beta, 1)$. The government introduces the following menu for agents of productivity $\theta_m : C_m = \{C^{TC}_m, C^{TI}_m\}$, where $C^{TC}_m$ consists of the persistent and deterrent allocations and $C^{TI}_m$ consists of the real and imaginary allocations.

**Definition 10** A direct betting mechanism for sophistication $\hat{\beta} \in (\beta, 1)$ with TC agents has $C = \{C^{TC}_m, C^{TI}_m\}_{\theta_m \in \Theta}$, with $C^{TC}_m = \{(c^{P}_m, y^{P}_m); (c^{D}_m, y^{D}_m)\}$ and $C^{TI}_m = \{(c^{R}_m, y^{R}_m); (c^{I}_m, y^{I}_m)\}$ satisfying: (i.) fooling constraints: $\forall \theta_m, \theta_m' \in \Theta, (c^{P}_m, y^{P}_m) \in C^{\hat{\beta}}_m$ and $(c^{D}_m, y^{D}_m) \in C^{\hat{\beta}}_m$, (ii.) deterrent constraints: $\forall \theta_m, \theta_m' \in \Theta, (c^{R}_m, y^{R}_m) \in C^{1}_{m}$, and $(c^{I}_m, y^{I}_m) \in C^{1}_{m}|_{m}$. 

26
The allocations that are truthfully implementable by a direct betting mechanism with TC agents is bounded by the incentive compatibility constraints (8) and (9) and the executability constraints (5). The definition of truthfully implementable allocations in this mechanism is similar to Definition 9.

**Definition 11** The allocation \( \{ (c^P_m, y^P_m), (c^R_m, y^R_m) \} \) is truthfully implementable for present bias \( \beta \) by a direct betting mechanism for sophistication \( \hat{\beta} \) with TC agents if there exists \( \{ (c^D_m, y^D_m), (c^I_m, y^I_m) \} \) such that (i.) incentive compatibility, and (ii.) executability are satisfied.

The betting mechanism with TC agents can lead to lower welfare than conditional commitment mechanisms. Notice that the threat allocations in the conditional commitment mechanism can be designed such that the TC agents would never choose it. However, it is difficult to deter TC agents from selecting the imaginary allocations in a betting mechanism.

To see this, let the persistent and real allocations be the ones implemented in an optimal conditional commitment mechanism. For them to be implemented in a betting mechanism, the incentive compatibility constraint for TC agents imply

\[
U_0(c^P_m, y^P_m; \theta_m) \geq U_0(c^I_m, y^I_m; \theta_m).
\]

Also, the TI agents at \( t = 0 \) would consume the real allocations, so (14) can be rewritten as

\[
\left[ u_0(c^P_{m,0}) - h_0\left(\frac{y^P_{m,0}}{\theta_m}\right) \right] - \left[ u_0(c^R_{m,0}) - h_0\left(\frac{y^R_{m,0}}{\theta_m}\right) \right] \geq U_1(c^I_m, y^I_m; \theta_m) - U_1(c^P_m, y^P_m; \theta_m).
\]

Similarly, the downward incentive compatibility constraint of the TI agents imply

\[
U_1(c^I_m, y^I_m; \theta_m) - U_1(c^I_{m-1}, y^I_{m-1}; \theta_m) \geq \frac{1}{\beta} \left( u_0(c^R_{m-1,0}) - h_0\left(\frac{y^R_{m-1,0}}{\theta_m}\right) \right) - \left[ u_0(c^R_{m,0}) - h_0\left(\frac{y^R_{m,0}}{\theta_m}\right) \right].
\]

Notice that the optimal allocations implemented in a conditional commitment mechanism might not be implementable in a betting mechanism, because there may not exist imaginary allocations that satisfy both TC and TI incentive compatibility constraints for low values of \( \beta \). As a result, when TC agents are present, conditional commitment mechanisms are more effective than betting mechanisms. This is because TC agents can mimic TI agents and select the imaginary allocations, which places an additional constraint on the imaginary

---

21 By Theorem 3, if \( \theta_{m-1} > \bar{\theta} \), then \( u_0(c^R_{m-1,0}) - h_0\left(\frac{y^R_{m-1,0}}{\bar{\theta}_m}\right) > u_0(c^R_{m,0}) - h_0\left(\frac{y^R_{m,0}}{\bar{\theta}_m}\right) \).
allocations. This has two critical implications. Firstly, this implies that conditional commitment mechanisms are more appealing than betting mechanisms. Secondly, since conditional commitment mechanisms have no effect on fully naïve agents, there is incentive for the government to help agents learn their present bias so they would be at least partially naïve. In section 7, I discuss how mechanisms that are robust to perturbations in retirement decisions are related to the setting with TC agents and fully naïve TI agents.

6 Social Security and Retirement Savings Accounts

In this section, decentralization of the optimal allocations in Sections 4 and 5 will be presented. I will first discuss the design of policies in an environment without TC agents. The betting mechanism can be decentralized with social security, and the conditional commitment mechanism can be decentralized with retirement savings accounts. A combination of both social security and retirement savings accounts can decentralize the hybrid mechanism. I will then discuss a decentralization with TC agents. This section also emphasizes the significant differences between the policy recommendations here and the ones that are being discussed in the literature and by policymakers.

6.1 The Timing of Claiming Social Security Benefits

A majority of the US population relies on social security benefits as their primary source of income during retirement.\(^{22}\) Also, while the US population is living longer, the average retirement age has remained steady for the past decade.\(^{23}\) This increases the duration of relying on social security benefits. Consequently, discussions on social security reforms to improve retirement welfare while maintaining its sustainability is an important policy issue.

A retiree in the US can choose when they wish to start claiming social security benefits. The earliest age possible for receiving benefits is 62. A person can delay claiming and receive higher monthly benefits for the rest of his/her life.\(^{24}\) Several papers have shown it is optimal for most people to delay benefits claiming, and that average Americans are receiving benefits too early (see Footnote 9). Knoll and Olsen (2014) find that the age of 62 is the most frequent

\(^{22}\)According to the Social Security Administration, nine out of ten individuals aged 65 or older receive social security benefits. Also, among the elderly beneficiaries, over half of the households receive over 50% or more of their income from social security.

\(^{23}\)See Munnell (2015).

\(^{24}\)For example, according to the Social Security Administration, the average monthly social security benefit for a beneficiary who started claiming at the age of 62 in 2014 is $1,098. If the same beneficiary waited till the age of 70 to start claiming benefits (the oldest enrollment age possible), then the monthly benefits would increase to $1932.
enrollment age, and the age of 70 to be the least frequent. As a result, early claimants are stuck with lower monthly benefits for the rest of their lives. 

Knoll et al. (2015) show that people expect to retire and claim benefits later, but many end up retiring and claiming benefits earlier than they have initially planned. This suggests that time inconsistency with present-bias could explain the tendency to claim early. It also suggests that people are non-sophisticated. Knoll et al. (2015) devise effective choice architectures to delay claiming.

I propose a new approach to this issue. Given the benefits structure, the labor decisions of the agents are made according to the benefits received later. However, agents claim their benefits earlier than planned due to present bias. In other words, whether the government knows it or not, the benefits for claiming late affects the pre-retirement labor decision of agents, while the benefits for claiming early affects the retirement consumption of agents. Consequently, optimal social security reforms should take time-inconsistent behavior and non-sophistication as given when designing the benefits. The government can vary the progressivity of social security benefits with the age of initial claiming to decentralize the betting mechanism. Benefits are progressive if the ratio of lifetime benefits to lifetime payroll taxes is higher for low income individuals than it is for people with higher average income. Benefits decrease in progressivity if the difference of this ratio between the low income and higher income individuals decreases.

With less progressive benefits for agents who claim late, non-sophisticated agents can be encouraged to work efficiently at a younger age. Unknowingly, non-sophisticated agents would want to claim benefits earlier than expected. Therefore, with more progressive benefits for early claimants, redistribution can be increased without distorting labor supply. In essence, the imaginary allocations are the benefits for claiming late, and the real allocations are the benefits for claiming early. The government and the agents bet on when the agents would claim their benefits and the progressivity is the wager.

To be more concrete, consider the following social security policy: the agents work in $t = 0$, and decide whether to claim benefits $b_1(y_0, y_1)$ in $t = 1$ or to claim $b_2(y_0, y_1)$ when they retire in $t = 2$. Social security as a policy with tax $T_t$ is defined as $P^{ss} = (b_1(y_0, y_1), b_2(y_0, y_1), T_0(y_0), T_1(y_0, y_1, k_1))$. The budget constraint in $t = 0$ is standard: $c_0 + k_1 \leq y_0 - T_0(y_0)$, where $k$ denotes savings. In $t = 1$, agents choose to claim early benefits $b_1(y_0, y_1)$ if $c_0 + c_1 + k_2 \leq y_1 - T_1(y_0, y_1, k_1)$. The government can vary the progressivity of social security benefits with the age of initial claiming to decentralize the betting mechanism.
at $t = 1$ or delay and claim at $t = 2$:

$$c_1 + k_2 \leq y_1 + k_1 + 1_t b_1 (y_0, y_1) - T_1 (y_0, y_1, k_1),$$

where $1_t$ is an indicator function that is equal to 1 if benefits are claimed at $t$ and zero otherwise. In $t = 2$, the agents face the following budget constraint: $c_2 \leq k_2 + 1_t b_1 (y_0, y_1) + 1_2 b_2 (y_0, y_1)$, so consumption in retirement depends on savings and benefits. Notice that if the agent started claiming at $t = 1$, then the benefit in $t = 2$ is $b_1$. This models the current social security system, where the benefits depend on the time it was initially collected and the amount would perpetuate till death. The following proposition demonstrates the decentralization in an environment with fixed present bias and sophistication.

**Proposition 3** If $\beta \in (\beta, 1]$, then the efficient allocation can be decentralized by $P^{ss}$, where $b_2 (y_0, y_1)$ is increasing and less progressive in income $(y_0, y_1)$ than $b_1 (y_0, y_1)$.

Proposition 3 shows how social security can implement the efficient allocation with non-sophisticated agents. The social security benefits $b_2$ is regressive in income to incentivize productive agents to produce efficiently. The benefits for early enrollment $b_1$ is a lump-sum transfer that provides full insurance and consumption smoothing for early retirees. Since the agents are non-sophisticated, they imagine claiming $b_2$ and would thus work efficiently in $t = 0$. However, the present-biased agents would claim $b_1$.

The current US system has benefits that are equally progressive in income for both early and late claimants. A reform along the lines proposed in Proposition 3 would make the benefits even more progressive for early claimants but less so for late claimants. Since the US population is already claiming earlier than planned, such a reform can help increase output efficiency, which would help increase taxable income and raise sustainability, while simultaneously improve social insurance. Both of which are goals stated in the National Commission on Fiscal Responsibility and Reform (2010).

### 6.2 Liquidity of Defined Contribution Plans

The design of defined contribution (DC) plans is of growing interest. The literature has focused on how to influence DC plan enrollment behavior. Other aspects of the design of DC plans has also gained attention. In particular, Beshears et al. (2015b) showed the DC

---

²⁹ Though most people choose to claim benefits during retirement, it is possible to claim benefits while working. For implementation, a penalty can be included for retiring early. One of the key recommendations of National Commission on Fiscal Responsibility and Reform (2010) is to raise the full retirement age.

plans in the US to be liquid: after separating from their employer, workers in the US can
move their DC account balance to an IRA or Roth IRA and withdraw for any reason from
the account before the eligibility age of 59.5 subject to a tax penalty of 10%. Such liquidation
before eligibility is forbidden in many countries, except under special circumstances. Compara-

tively, DC plans in the US are flexible and meet the transitory needs of a worker. However, flexibility is undesirable if early withdrawal is due to present bias.

Beshears et al. (2015a) showed that making the DC plan more illiquid for time-
inconsistent agents can be an effective commitment device and increase savings. This paper
provides an alternative view: such commitment can be provided in exchange for efficient out-
put. The liquidity of DC plans can be redesigned to depend on income and act as a threat
to sophisticated time-inconsistent agents, which incentivize agents to produce efficiently and
allow the government to provide better insurance. The conditional commitment mechanism
can be decentralized with early withdrawal as the off-equilibrium threat. This idea is similar
to the implementation of partial illiquidity in Bond and Sigurdsson (2017), where the agent
in $t = 1$ can increase $c_1$ and decrease $c_2$ to punish misbehavior in $t = 0$.

To implement full efficiency with retirement accounts, consider the following timing:
agents are endowed with $s_0 > 0$ in the accounts and work and deposit $s_{t+1}$ into their
accounts in $t$, and are allowed to withdraw $\eta \in (0, s_0)$ early from their accounts and
receive $\xi (y_0, s_1) - \eta$ as a subsidy in $t = 1$.\footnote{For simplicity, I do not consider liquid savings accounts, like bank savings.} Let $\tau (y_0, y_1, s_1)$ be the early withdrawal
penalty, which is an off-equilibrium path threat contingent on income history. A retirement account with contemporaneous income tax $T_t$ and savings subsidy $\rho_t$ is defined as
$P_{ra} = (s_0, \tau (y_0, y_1, s_1), \xi (y_0, s_1), \eta, \rho_1, \rho_2, T_0 (y_0), T_1 (y_1))$.

In $t = 0$, the budget constraint is: $c_0 + \frac{s_2}{1 + \rho_2} \leq y_0 - T_0 (y_0)$. In $t = 1$, agents choose whether to withdraw early from the retirement account:

$$c_1 + \frac{s_2}{1 + \rho_2} \leq 1_{EW} \xi (y_0, s_1) + y_1 - T_1 (y_1),$$

where $1_{EW}$ is equal to 1 if the agent withdrew early and zero otherwise. In $t = 2$, agents
face the following budget constraint:

$$c_2 \leq 1_{EW} (1 - \tau (y_0, y_1, s_1)) (s_0 + s_1 + s_2 - \eta) + (1 - 1_{EW}) (s_0 + s_1 + s_2). \tag{15}$$

\footnote{For example, countries such as Germany, Singapore and the UK.}
\footnote{Argento, Bryant, and Sabelhaus (2015) find 45% of contributions to retirement accounts among participants under the age of 55 in 2010 were offset by early withdrawals, which is higher than years prior to the Great Recession. Munnell and Webb (2015) estimates that if early withdrawals were not possible total 401(k) wealth would be 25% higher and total IRA wealth would be 23% higher.}
\footnote{For example, see Ashraf, Karlan, and Yin (2006) and Beshears et al. (2015a).}
Inequality (15) shows that, with early withdrawal, $c_2$ decreases proportionally to the penalty $\tau(y_0, y_1, s_1)$, where

$$\tau(y_0, y_1, s_1) = \begin{cases} \hat{\rho}(y_0, s_1) & \text{if } y_1 \geq \bar{y}(y_0, s_1) \\ 1 & \text{if } y_1 < \bar{y}(y_0, s_1) \end{cases}.$$  

The penalty is structured so that if $y_1$ is commensurate with $y_0$ (i.e. $y_1 < \bar{y}(y_0, s_1)$), then the agent would not be tempted to withdraw from the retirement account (because, if he does withdraw early, then $c_2 = 0$). Therefore, the retirement account is illiquid and the agent is committed to having sufficient savings for retirement. If $y_1$ is large (at least $\bar{y}(y_1, s_1)$), then the present-biased agent will withdraw early and be penalized by having lower retirement savings. Only agents who produced an inefficiently low output in $t = 0$ with respect to their productivity would be tempted to withdraw early in $t = 1$. Consequently, agents would produce efficiently in $t = 0$ to avoid the temptation of withdrawing early in $t = 1$. The following proposition demonstrates the decentralization in an environment with fixed present bias and sophistication.

**Proposition 4** If $\hat{\beta} \in [\beta, 1)$, then the efficient allocation can be decentralized by $P^{ra}$.

There are proposals to make the account more liquid the lower the income, which is consistent with Proposition 4. However, there are two main differences between current plans and the plan proposed in Proposition 4. First, liquidity in the proposed retirement plan is not indiscriminately available for all low income agents. To be able to withdraw early, an agent must earn a sufficiently larger income than the previous period to qualify. Another difference is that in Proposition 4, the early withdrawal penalty tax $\tau$ is applied to the residual amount left in the savings account. This penalty tax decreases the savings available in the retirement account to discipline the younger self. The current system has the early withdrawal penalty tax on the withdrawal amount. Though the current system discourages early withdrawals and helps smooth consumption, it does not have the disciplining effect on the younger-self to encourage efficient output.

### 6.3 Decentralizing the Hybrid Mechanism

By Section 4, the efficient allocation is implementable by a hybrid mechanism in an environment where all agents are time-inconsistent with heterogeneous sophistication and

---

35 Obama proposed a hardship exception to early withdrawal penalties for those who have received unemployment for more than 26 weeks, see Fiscal Year 2017 Budget (2016).  
36 It is also possible to implement this as a repayment scheme where the agents are required to replenish their accounts by a large amount if they withdrew early.
present bias. The hybrid mechanism can be decentralized with social security and retirement savings accounts.

The government endows the agents an initial savings of $s_0 > 0$ in their retirement accounts at the beginning of $t = 0$. The agents deposit $s_t$ in their DC accounts, which maintain the same penalty features introduced previously. The social security benefits have the original structure in terms of progressivity, but the benefits are decreased by $s_0$. The government designs both $P^{ss}$ and $P^{ra}$ for some sophistication $\hat{\beta} \in \left(\beta, 1\right)$. $P^{ss}$ is designed for the least present-biased agents $\beta$, and $P^{ra}$ is designed for the most present-biased agents $\beta$. The initial savings $s_0$ is necessary, because the threat of a liquid retirement account is credible only if there were funds in the account. If $s_0 = 0$, sophisticated agents can always work inefficiently, choose a low $s_1$ and claim social security benefits early to mimic naïve agents.

**Proposition 5** For the environment where all agents are time-inconsistent with hidden present bias and sophistication, the efficient allocation $\{(c^*_m, y^*_m)\}_{\theta_m \in \Theta}$ can be decentralized by $P^{ss}$ and $P^{ra}$.

Proposition 5 follows from Propositions 3 and 4 and the proof of Theorem 1. An example of such an implementation is the central provident fund (CPF), a retirement account in Singapore. All conscripts are endowed with at least 5000 SGD in their CPF accounts after military service. Since national service is compulsory for all males in Singapore, this endowment is similar to the initial endowment of $s_0 > 0$. However, early withdrawal from the CPF account is not allowed.

It should be noted that Moser and de Souza e Silva (2017) also provide a similar implementation albeit in a different environment. They consider an environment with dynamically stochastic present bias, so off-equilibrium path policies are ineffective. As a result, they arrive at a different conclusion on the design of social security and retirement accounts. Information rents are provided to more productive agents in the form of flexible savings plans like a defined contribution plan, while social security is less flexible and is for less productive agents. This is different from the implementation in this paper, where social security and retirement accounts are used to separate sophistication and present bias.

### 6.4 Retirement Savings Policies with Time-Consistent Agents

From Section 5, the key insight in the mechanism with TC agents is the necessity of deterrent allocations. I will demonstrate an implementation where the decision to enroll

---

37 Singaporeans are allowed to withdraw money from their CPF accounts for housing and education, but must repay it with interest. Medical expenses are exempt from repayment.
in retirement savings accounts is voluntary. More specifically, TI agents are encouraged to enroll in the program, which provides them with commitment. TC agents are encouraged to opt out of the program and save on their own accord. This is because TC agents do not need the commitment provided by the retirement savings accounts, while full discretion in savings is undesirable for TI agents.

More specifically, to implement the optimal allocation presented in Section 5, the government makes participation in retirement savings accounts a voluntary decision. The retirement savings accounts, $P_{\text{ra}}$, takes the form described in Proposition 4. Agents have a choice of enrolling in $P_{\text{ra}}$ or save on their own at interest rate $r = \frac{1}{\delta} - 1 = 0$. Income taxes would depend on the enrollment decision of the agents. Let $T_{\text{out}}$ denote the income tax for agents who did not enroll in $P_{\text{ra}}$. Furthermore, to deter less present-biased agents from opting out, it may be necessary to introduce consumption subsidies conditional on savings, $\tau_{1,\text{out,ns}}$, and after-tax income $I_{1,\text{out,ns}}$ in $t = 1$ for those who chose to save on their own, which is triggered when intertemporal wedge in $t = 1$ is not zero: $u_1'(c_1) \neq u_2'(c_2)$.\[38]

**Proposition 6** For the environment with TC agents, the optimal allocation can be decentralized by $\{\tau_{\text{out}}, T_{\text{out}}, P_{\text{ra}}\}$.

Proposition 6 leads us to two important insights. Firstly, not only should the retirement savings accounts be partially liquid, its participation should also be voluntary. Secondly, when agents are not fully na"ive, with a properly designed partially liquid retirement savings account in place, the government does not need to provide additional paternalistic measures to help agents save more for retirement. In fact, TI agents should be punished with low retirement savings and high present consumption when they do not participate in retirement savings accounts.

### 6.5 Back-of-the-Envelope Quantitative Analysis

### 7 Discussion

In this section, I will discuss some limitations of the results from the previous sections and some possible extensions.

\[38\] For agents with sufficiently small $\beta$, the lack of commitment from saving on their own is a strong enough incentive to participate in retirement savings accounts.
7.1 Limited Promises and Punishments

Assumption 1 provides a non-empty set of bets that could deceive non-sophisticated agents for any information rent. If Assumption 1 fails, the efficient optimum might not be implementable. To see this, consider the case where the utility function is bounded below. Figure 5 illustrates how betting can be limited for fully naïve agents with $\theta \in \{\theta_L, \theta_H\}$. The flatter solid (blue) curve represents the indifference curve of the ex-ante utility and the steeper solid (red) curve represents the indifference curve of the ex-post utility, both evaluated at allocation $(c_1^*, c_2^*)$. The dotted (blue) curve indicates the minimum information rent necessary for the productive agents to be truthful. However, the best the government can do is to set the imaginary allocation at the boundary as indicated in Figure 5.

As a result, asymmetric information causes distortions if the imaginary allocations cannot fully cover the minimal information rent necessary for truth-telling. The government would have to distort consumption and output downwards for lower types, which can be implemented by income taxation as in Mirrlees (1971). A similar argument can be made for conditional commitment mechanisms. Figure 6 illustrates how threats are limited in a sophisticated case with two productivity types.

Nevertheless, the government is still able to improve welfare above the constrained efficient optimum by using betting or conditional commitment mechanisms when Assumption 1 fails. This is because a portion of information rents is loaded on the off-equilibrium path allocations, which helps relax incentive compatibility and decrease distortions.
7.2 Trembling Hand

The betting and conditional commitment mechanisms introduce off-equilibrium path allocations to relax the incentive compatibility constraint. A legitimate concern when decentralizing these mechanisms is the potential for agents to unintentionally select the off-equilibrium allocations. This could lead to significant welfare loss for the agent. For example, by introducing liquid retirement accounts as prescribed in Proposition 4, agents could accidentally withdraw early from the retirement savings account. It could also be costly for the government. For example, agents could accidentally claim their social security benefits late. With the higher and less progressive benefits for late claimants prescribed in Proposition 3, this could lead to large revenue losses for the government.

For illustrative purposes, I will focus the discussion on a fixed present bias $\beta$ and sophistication $\hat{\beta}$. Also assume that $u_t = u$ and $h_t = h$ for all $t$. Consider a betting mechanism where agents might unintentionally choose the imaginary allocation with probability $\kappa \in (0, 1)$. As a result, the imaginary allocations would enter the feasibility constraint. Assume that the agents are oblivious to the possibility of making a mistake with probability $1 - \kappa$ in $t = 1$, which is consistent with their incorrect belief of $\beta$. The following theorem characterizes the intertemporal wedge in this environment.

**Proposition 7** If $\kappa \in (0, 1)$, the intertemporal wedge has the following properties: (i.) $\tau_C^1 < 0$ for the imaginary allocations with $\theta_m > \theta_1$, and (ii.) $\tau_C^1 > 0$ for the real allocations with $\theta_m > \theta_1$.

Proposition 7 shows how consumption smoothing is not implementable for agents with $\theta_m > \theta_1$, and the best the government can do is to provide consumption smoothing for them in expectation. Since the imaginary allocation attempts to load the information rent on retirement consumption, those who choose it would over-save. Therefore, the agents choosing the real allocations would under-save. This is true even for the most productive agents, so there will be distortions at the top.

When $\hat{\beta} = 1$, Proposition 7 also characterizes the intertemporal wedge in an environment with both fully naïve TI agents and a mass of $1 - \kappa$ TC agents. The government is unable to separate consistency, because TI agents believe themselves to be TC. As a result, deterrent allocations are useless and the TC agents would consume the imaginary allocations on-path.

On the other hand, for conditional commitment mechanisms, if agents might choose the threat allocation, consumption smoothing is not implementable for lower productivity types. This is because the threat is introduced into the menu of low productivity types. Also, since the threat allocation works by exacerbating present bias, the threat allocation would have agents under-saving so the real allocation would have agents over-saving.
7.3 Outside Commitment Devices

In reality, self-control problems can be mitigated by a wide array of commitment devices available in the market. In the case of sophisticated agents, if commitment devices are available and its usage is unobservable, then threats are less potent. This is because agents can purchase commitment and bind themselves to an intertemporal allocation. Therefore, screening of productivity would be more costly when commitment devices for sophisticated or partially naïve agents are available. However, for non-sophisticated agents, the government can always choose imaginary allocations that make buying an outside commitment device undesirable. This has the additional benefit of preventing the non-sophisticated agents from using an inefficient amount of commitment (Heidhues and Koszegi 2009).

7.4 Paranoid Time-Consistent Agents

Previous analysis assumed the TC agents were sophisticated (\(\hat{\beta} = \beta = 1\)). A paranoid agent is a non-sophisticated TC agent who believes the future-self is present biased. Paranoia affects the behavior of TC agents, and consequently government policy can be adjusted to exploit it. Paranoid agents respond to threats, and can also be fooled. It is easy to see how a conditional commitment mechanism can extract information rents from paranoid agents, since the government can construct the threat allocation using similar methods for the TI agents. The betting mechanism is more subtle.

To see how a betting mechanism achieves the efficient optimum in an economy with only paranoid agents, \(\hat{\beta} < \beta = 1\), consider the example with \(\Theta = \{\theta_L, \theta_H\}\). Let \(C_m = \{(c^*, y^*_m), (c^*_m, y^*_m)\}\) with \(c^*_{m,0} = c^*_0\) and \(c^*_m = c^*\). The allocations satisfy the fooling constraints \(u_1(c^*_L,1) + \hat{\beta}u_2(c^*_L,2) \geq u_1(c^*_1) + \hat{\beta}u_2(c^*_2)\), the executability constraints \(u_1(c^*_1) + u_2(c^*_2) \geq u_1(c^*_L,1) + u_2(c^*_L,2)\), and the incentive compatibility constraints, which implies the following

\[
\sum_{t=0}^{1} h_t \left( \frac{y^*_H,t}{\theta_L} - h_t \left( \frac{y^*_L,t}{\theta_L} \right) \right) \geq [u_1(c^*_1)+u_2(c^*_2)] - [u_1(c^*_L,1)+u_2(c^*_L,2)] \geq \sum_{t=0}^{1} h_t \left( \frac{y^*_H,t}{\theta_H} - h_t \left( \frac{y^*_L,t}{\theta_H} \right) \right).
\]

The fooling and executability constraints imply \(c^*_m,1 > c^*_1\) and \(c^*_m,2 < c^*_2\). Combined with the incentive compatibility constraints, it must be that \(c^*_L,1 > c^*_H,1\) with \(c^*_L,2 < c^*_H,2\). In essence, the government fools the paranoid agents by choosing the imaginary allocations to exacerbate their fears. A paranoid agent would predict choosing the imaginary allocations even though the agent is strictly worse off by choosing it, because he/she does not think

---

39 There is a growing market for commitment devices. For example, StickK, Pact and Beeminder are some recent websites that offer commitment contracts.
the real allocation is attainable. The government takes advantage of this by making the imaginary allocation for the \( \theta_L \) agent even worse. Hence, the paranoid \( \theta_H \) agent produces efficiently because there is a fear that by misreporting, he/she would have even less savings.

The way the betting mechanism works for paranoid agents is in stark contrast to the logic presented in the previous sections. Non-sophisticated TI agents are fooled by empty promises, but paranoid agents are fooled by empty threats.

If the economy has both paranoid and TI agents, then using a betting mechanism could be problematic. For example, imagine an economy with paranoid TC agents with incorrect belief \( \hat{\beta} < 1 \), which corresponds to the belief of the non-sophisticated TI agents with \( \beta < \hat{\beta} \). If the government tries to fool the agents, then it is not possible for it to separate agents along consistency level. In this particular case, depending on who the government chooses to fool, either the paranoid TC agents would end up selecting the imaginary allocation used to fool the TI agents or the TI agent would choose the imaginary allocation used to fool the TC agents. The resulting welfare would be lower compared to when TC agents are sophisticated. However, such a problem does not arise when the government uses a conditional commitment mechanism, because the same threats for TI agents can also deter paranoid agents from misreporting.\(^{40}\)

### 7.5 Dynamic Stochastic Shocks

The recent literature on optimal dynamic Mirrlees taxation have focused on stochastically evolving productivity.\(^{41}\) To relax the incentive compatibility constraints, intertemporal distortions are introduced and are characterized by the inverse Euler equation \((\text{Golosov, Kocherlakota, and Werning, 2003})\). This is in contrast to the current paper, which assumes a constant productivity. This assumption seems innocuous since \(\text{Keane and Wolpin, 1997}\) found that the bulk of labor-market uncertainty can be explained by skill endowments in adolescence. Also, \(\text{Guvenen et al., 2016}\) showed that most individuals experience very little income change in a given year, which suggests a highly persistent income process. However, \(\text{Guvenen et al., 2016}\) also showed that the distribution of earnings changes exhibit high kurtosis, so a non-negligible number of people experience large income changes. Therefore, a characterization of optimal betting and conditional commitment mechanisms with dynamic stochastic productivity shocks would be useful for the design of policy.

Conditional commitment mechanisms could help raise welfare in settings with dynamic

---

\(^{40}\)This is true because the credible threat constraints and the incentive compatibility constraints are the same for both the TI and TC agents who share the same beliefs. Finally, the executability constraint is more relaxed for the TC agents than the TI agents.

\(^{41}\)See surveys \(\text{Golosov, Tsyvinski, and Werning, 2006}\) and \(\text{Kocherlakota, 2010}\) for a comprehensive list of papers.
shocks and time-inconsistent agents. Halac and Yared (2014) focus on public strategies, so players can only report their current taste shock. However, they point out that by allowing the players to report on past shocks, it can potentially relax the incentive compatibility constraint by punishing inconsistent reports. However, they show that this would not work in a setting with independent shocks as in Amador, Werning, and Angeletos (2006), so conditional commitment mechanisms are useful as long as shocks are persistent.

To see how, consider two productivity types, \( \Theta = \{ \theta_L, \theta_H \} \), and suppose Assumption 1 holds. If productivity shocks are independent across time, then \( \theta_0 \) provides no information on \( \theta_1 \), so off-equilibrium threats do not deter misreports. However, if shocks are persistent, then \( \theta_0 \) is informative of \( \theta_1 \). In particular, if \( \Pr(\theta_{L,1} | \theta_{L,0}) = 1 \) and \( \Pr(\theta_{H,1} | \theta_{H,0}) > 0 \), then the efficient allocation is implementable. The low productivity shocks need to be fully persistent, because low productivity agents in \( t = 0 \) need to be protected from the off-equilibrium punishment in \( t = 1 \), which is similar to Proposition 3 in Bond and Sigurdsson (2017).

Recent studies have documented the temporal stability of time preferences (Wolbert and Riedl, 2013; Meier and Sprenger, 2015; Dean and Sautmann, 2016). However, there is still evidence of temporal instability (see Meier and Sprenger (2015)). Moser and de Souza e Silva (2017) analyzed an environment where present bias is independently stochastic over time. They show that in their environment off-path policies do not relax the incentive compatibility constraints. Indeed, off-equilibrium path threats require the agents knowing what their future-selves might do and independently stochastic time inconsistency hinders the ability of agents to predict.

8 Summary and Conclusion

This paper provided methods on utilizing the agents’ time inconsistency to increase welfare above the constrained efficient optimum, contrary to traditional policy proposals, where the primary goal was to mitigate the present bias. These methods provide new insights on the progressivity of social security benefits and the liquidity of defined contribution plans.

The results of this paper could be applied to other settings, like the design of health or life insurance policies. The concept of betting and provision of commitment could potentially be used in a wider array of mechanism design problems with agents suffering from other biases, such as overconfidence.

Though welfare increases with the proportion of time-inconsistent agents in the economy, this paper does not advocate time-inconsistent behavior. The focus on savings has obscured other costs associated with being time-inconsistent, such as inadequate human capital development. Future work should explore this trade-off and its consequences on policy.
A Proofs

Proof of Proposition 3: Let \((c^l_m, y^l_m)\) be imaginary allocations constructed in a direct mechanism that support the efficient allocation, where \(y^l_m = y^*_m\) and \(c^l_{m,0} = c^*_{m,0}\) for all \(\theta_m \in \Theta\). Let \(Y^*_t = \{y^*_1, \ldots, y^*_m, \ldots, y^*_M, t\}\). Furthermore, if \(y_t \in Y^*_t\), let \(y_t^{-1} \in \Theta\) denote the corresponding type. For example, \(y_t^{-1} = \theta_m\). Let \(EO = \{y = (y_0, y_1) | y_0 \in Y^*_0, y_1 \in Y^*_1\} \) denote set of efficient output history. Consider the following taxes: \(T_0(y_0) = y_0 - c^*_m\), and \(T_1(y_0, y_1, k_1) = y_1 + k_1 + 1 \cdot b_1(y_0, y_1) + 1_2 \alpha (y_0, y_1) - c^*_1\), and benefits:

\[
\begin{align*}
    b_1(y_0, y_1) &= \begin{cases} 
    c^*_2 & \text{if } y \in EO \\
    0 & \text{otherwise}
    \end{cases}, \\
    b_2(y_0, y_1) &= \begin{cases} 
    c^l_{m,2} & \text{if } y = y^*_m \\
    0 & \text{otherwise}
    \end{cases},
\end{align*}
\]

with \(\alpha (y_0, y_1) = c^*_1 - c^l_{m,1}\), if \(y \in EO\) and \(y_t^{-1} = \theta_m\), otherwise \(\alpha (y_0, y_1) = c^*_1\). This construction implements the efficient allocation.

Also, \(b_2\) can be constructed to be less progressive than \(b_1\). To see how, note that from the proof of Theorem 1 by Assumption 1, it is always possible to lower \(c^l_{m,1}\) and increase \(c^l_{m,2}\) such that incentive compatibility holds. Since \(u_t\) is strictly concave, the increase in \(c^l_{m,2}\) would need to be large compared to the decrease in \(c^l_{m,1}\). This increases the ratio of lifetime benefits to taxes paid for higher productivity agents claiming at \(t = 2\), which decreases progressivity.

Proof of Proposition 4: First, consider on-equilibrium path policies (\(1_{EW} = 0\)). Set \(s_0 \in (0, c^*_2), \rho_1 = -1\) and \(\rho_2 = \frac{1}{\beta} - 1\), so agents do not save at \(t = 0\). When \(1_{EW} = 0\), income taxes are

\[
    T_0(y_0) = \begin{cases} y_0 - c^*_0 & \text{if } y_0 \geq y^*_{1,0} \\
    y_0 & \text{otherwise}
    \end{cases}, \quad T_1(y_1) = \begin{cases} y_1 - c^*_1 - \beta (c^*_2 - s_0) & \text{if } y_1 \geq y^*_{1,1} \\
    y_1 & \text{otherwise}
    \end{cases},
\]

With this setup, agents in \(t = 0\) and \(t = 1\) would choose the efficient consumption on-equilibrium path.

Next, consider off-equilibrium path policies to support efficient output. A \(\theta_{m+1}\) agent with \(y_0 = y^*_{m,0}\) predicts the future-self would solve the following in \(t = 1\) if \(y_1 < \bar{y} (y^*_{m,0})\):

\[
    \max_{c_1, c_2, y_1} u_1 (c_1) - h_1 \left(\frac{y_1}{\theta_{m+1}}\right) + \bar{\beta} u_2 (c_2) \quad \text{subject to } c_1 + \beta (c_2 - s_0) = y_1 - T_1 (y_1).
\]

Let \((\tilde{c}_{m,1}, \tilde{c}_{m,2}, \tilde{y}_{m,1})\) denote the solution. Notice the solution to the problem does not depend on \(\theta_m\). If \(y_1 = \bar{y} (y^*_{m,0})\), the agent in \(t = 1\) is predicted to solve the following:
\[
\max_{c_1,c_2} u_1(c_1) + \hat{\beta}u_2(c_2) \text{ subject to } c_1 + \beta \left( \frac{c_2}{1-\rho(y^*_{m,0})} - s_0 + \eta \right) = \xi(y_0) + y_1 - T_1(y_1),
\]
and let \((\hat{c}_{m,1}, \hat{c}_{m,2})\) denote the solution, which does not depend on \(\theta_m\) nor \(\theta_{m+1}\).

Let \(\bar{y}(y^*_{m,0}) = y^*_{m,1} + \alpha_m\), with \(\alpha_m > 0\). Set \(\hat{\rho}(y^*_{m,0}) > 1 - \beta\) to tempt the agent to increase \(c_1\) and decrease \(c_2\), and construct \(\xi(y^*_{m,0}) > 0\) to ensure the threat constraint binds. Let \(\xi_m(\hat{\rho}(y^*_{m,0}), \alpha_m)\) denote this for \(y_0 = y^*_{m,0}\). Hence, types with \(\theta > \theta_{m+1}\) would strictly prefer withdrawing early if \(y_0 = y^*_{m,0}\).

Next, choose \(\hat{\rho}(y^*_{m,0})\) to satisfy the local downward incentive compatibility constraint of type \(\theta_{m+1}\). Notice that, for a given \(\alpha_m > 0\), \(\hat{c}_{m,2}(\hat{c}_{m,1})\) is a decreasing (increasing) function of \(\hat{\rho}(y^*_{m,0})\). The government can then increase \(\hat{\rho}(y^*_{m,0})\) till the incentive compatibility constraint holds, for any given \(\alpha_m > 0\). Let \(\hat{\rho}(\alpha_m)\) denote the level of early withdrawal penalty such that the incentive compatibility constraint holds. Hence, the government can set \(\xi(\hat{\rho}_m(y^*_{m,0}), \alpha_m)\) and \(\hat{\rho}(y^*_{m,0}, \alpha_m)\) for the threat and local downward incentive compatibility constraints to hold for type \(\theta_m\).

Finally, to pin down \(\alpha_m\), the government can increase \(\alpha_m\) till the executability constraint for \(\theta_m\) holds. Also, since the threat constraint binds, the \(\theta_m\) agent would predict an output of \(y < \bar{y}(y^*_{m,0})\) so the credible threat constraints for \(\theta_m\) hold. The process above can be repeated for all productivity types to ensure global incentive compatibility.

**Proof of Proposition 6**: Following the same process as in the proof of Proposition 4, let \(P^\theta\) be chosen to implement the real allocations \((c^R_m, y^R_m)\) for all \(\theta_m \in \Theta\).

Next, let \(Y^P_t = \{y^P_{1,t}, \ldots, y^P_{m,t}, \ldots, y^P_{M,t}\}\). Furthermore, if \(y_t \in Y^P_t\), let \(y_t^{-1} \in \Theta\) denote the corresponding type. Let \(EO^{out}_t = \{y = (y_0, y_1) | y_0 \in Y^P_0, y_1 \in Y^P_1 \text{ and } y_0^{-1} = y_1^{-1}\}\) denote the set of equilibrium output history. For all \(t < 2\), let \(\hat{c}_t : \mathbb{R}_+ \mapsto \mathbb{R}_+\) be such that

\[
\begin{align*}
\hat{c}_t(y_t) &\equiv \max_{y'_t \leq y_t} \hat{c}(y'_t) \text{ subject to } y'_t \in Y^P_t \text{ and } \hat{c}(y^P_{m,t}) = c^P_m. \\
\end{align*}
\]

Similarly, for all \(t < 2\), and \(\hat{c}_2 : \mathbb{R}_+ \mapsto \mathbb{R}_+\) define \(\hat{c}_2(y_t) \equiv \max_{y'_t \leq y_t} \hat{c}(y'_t) \text{ subject to } y'_t \in Y^P_t \text{ and } \hat{c}_2(y^P_{m,1}) = c^P_m.\) Also, choose \(T^\text{out}_t\) as a function of output history:

\[
T^\text{out}_0(y_0) = \begin{cases} 
\frac{y_0 - \hat{c}_0(y_0)}{y_0} & \text{if } y_0 \geq y^P_{1,0} \\
\frac{y_0}{y_0} & \text{otherwise}
\end{cases}, \\
T^\text{out}_1(y_0, y_1) = \begin{cases} 
\frac{y_1 - \hat{c}_1(y_1) - \hat{c}_2(y_1)}{y_1} & \text{if } y \in EO^{out}_1 \\
y_1 & \text{otherwise}
\end{cases}.
\]

By the construction of \(\hat{c}_t\) and \(T^\text{out}_t\), agents do not have an incentive to produce outside of the set \(Y^P_t\) for all \(t\). Furthermore, by the construction of \(T^\text{out}_1\), agents of productivity \(\theta_m\) would produce \(y^P_m\), since after-tax income would be zero if \(y \not\in EO^{out}\). Since the persistent allocations are incentive compatible, TC agents would produce according to their productivity.
Finally, let $c^D_1$ and $c^D_2$ satisfy (12) and (13). To separate TI agents from TC agents, choose $1 + \tau_1^{out,ns} = \frac{u_1'(c^D_1)}{\beta u_2'(c^D_2)}$ and $I^{out,ns} = (1 + \tau_1^{out,ns}) c^D_1 + c^D_2$. As a result, a TI agent mimicking a TC agent would predict consuming $(c^D_1, c^D_2)$ when solving: \[ \max_{c_1, c_2} u_1(c_1) + \hat{\beta} u_2(c_2) \text{ subject to } (1 + \tau_1^{out,ns}) c_1 + c_2 \leq I^{out,ns}, \] which is triggered whenever $u_1'(c_1) \neq u_2'(c_2)$.

References


